Observatoire ARGA

МЕЖДУНАРОДНЫЙ АНАЛИТИЧЕСКИЙ ДОКЛАД

Исследование №20

Международная модель независимого AI Observatory

Общественный мониторинг, система показателей, реестр рисков и институциональная независимость

Организация: Observatoire ARGA

Русская версия

21 rue de l’Aviation, 64600 Anglet, France

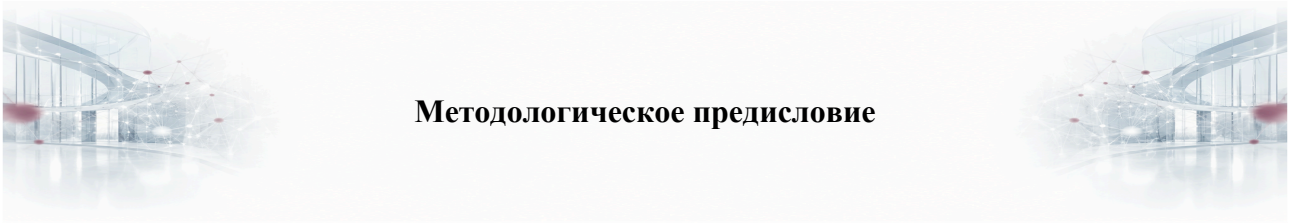
info@argaobservatory.org | www.argaobservatory.org

22 октября 2026 года

Содержание

Методологическое предисловие — 3
Научные основания и исследовательский дизайн — 4
1 Предмет исследования и исследовательский вопрос — 6
2 Определение независимого AI Observatory — 7
3 Международные институциональные ориентиры — 9
4 Правовая природа мониторинга и юридические границы — 11
5 Объекты наблюдения и реестр систем — 13
6 Организационная архитектура — 14
7 Институциональная независимость — 16
8 Финансирование и противодействие захвату повестки — 18
9 Конфликты интересов и отводы — 20
10 География и сопоставимость юрисдикций — 21
11 Стратегия общественного мониторинга — 22
12 Приём сообщений и защита заявителей — 24
13 Классификация событий: инцидент, опасность, сигнал — 25
14 Система показателей: принципы — 27
15 Знаменатель, отбор и смещения наблюдений — 29
Изменение числа событий при расширении источников — 29
16 Сбор, происхождение и качество доказательств — 31
17 Проверка достоверности информации и право на ответ — 32
18 Реестр рисков: структура и смысл — 34
19 Шкалы риска и правила приоритизации — 35
20 От мониторинга к рекомендациям и защите прав — 36
21 Защита персональных данных и информационная безопасность — 38
22 Уязвимые группы и участие пострадавших — 40
23 Международный обмен данными и трансграничные кейсы — 41
24 Алгоритмы в санкционном комплаенсе — 42
25 Международный розыск, идентификация и миграция — 44
26 Финансовый мониторинг AML/KYC — 46
27 Государственные услуги и автоматизация решений — 47
28 Закупки ИИ и контракты публичного сектора — 48
29 Frontier AI и модели общего назначения — 50
30 Демократические процессы и информационная целостность — 51
31 Экология, труд и общественные последствия — 52
32 Взаимодействие с заявителями и гражданским обществом — 54
33 Процедуры публикации и категории материалов — 55
34 Управление исправлениями, отзывами и оспариванием — 57
35 Контроль качества исследований и peer review — 59
Разбор согласованности независимого кодирования — 60
36 Цифровая инфраструктура и технический контроль — 62
37 Метрики деятельности самой обсерватории — 63
38 Сотрудничество с международными организациями — 64
39 Пилотная программа Observatoire ARGA — 66
40 Финансовая и кадровая реализуемость — 67
41 Выводы и предлагаемая международная модель — 69
Практические приложения: рабочие формы и условные сценарии — 70
Приложение 1 Устав независимого AI Observatory — 70
Приложение 2 Хартия независимости и редакционных гарантий — 71
Приложение 3 Декларация конфликта интересов — 72
Приложение 4 Реестр финансирования и связанных услуг — 73
Приложение 5 Матрица компетенций и назначений — 74
Приложение 6 Паспорт наблюдаемой AI-системы — 75
Приложение 7 Карточка сообщения об AI-инциденте — 75
Приложение 8 Регистр AI-hazards и предварительных сигналов — 76
Приложение 9 Реестр рисков и контрольных мер — 77
Приложение 10 Словарь ключевых показателей — 78

Приложение 11 Паспорт статистического расчёта — 78
Приложение 12 Классификатор качества источников — 79
Приложение 13 Журнал цепочки хранения доказательств — 80
Приложение 14 Протокол независимой проверки карточки — 81
Приложение 15 Процедура защиты чувствительного заявителя — 81
Приложение 16 Стандарт запроса о комментарии и праве на ответ — 82
Приложение 17 Порядок исправления, отзыва и замены отчёта — 83
Приложение 18 Матрица применимости права по юрисдикции — 84
Приложение 19 Страновой профиль AI governance — 85
Приложение 20 Паспорт методики международного сопоставления — 86
Приложение 21 Методика отраслевого скрининга AML и санкций — 87
Приложение 22 Протокол проверки AI-систем в международном розыске — 88
Приложение 23 Матрица человеческого контроля и обжалования — 89
Приложение 24 Форма отчёта о внешней рецензии — 89
Приложение 25 Редакционный паспорт публикации — 90
Приложение 26 Шаблон обращения к регулятору или публичному органу — 91
Приложение 27 Протокол срочной аналитической эскалации — 92
Приложение 28 Правила визуализации публичного dashboard — 92
Приложение 29 План пилота на двенадцать месяцев — 93
Приложение 30 Сценарий: ложное санкционное совпадение — 94
Приложение 31 Сценарий: риск ошибочной идентификации на границе — 95
Приложение 32 Сценарий: социальная выплата и непрозрачный скоринг — 95
Приложение 33 Сценарий: обновление модели без повторной оценки — 96
Приложение 34 Сценарий: журналистский всплеск AI-инцидентов — 97
Приложение 35 Паспорт ежегодного отчёта обсерватории — 98
Приложение 36 Сводная матрица принятия решения о запуске — 99
Приложение 37 Схема данных и связей между исследовательскими реестрами — 100
Приложение 38 Согласованность классификации независимыми оценщиками — 100
Приложение 39 Проверка временных рядов и структурных разрывов — 101
Приложение 40 Расширенная матрица оценивания общественного вреда — 102
Приложение 41 Запрос поставщику ИИ о верифицируемости системы — 103
Приложение 42 Анализ чувствительности показателя с условным примером — 103
Приложение 43 Порядок обработки запроса об удалении или ограничении данных — 104
Приложение 44 Модель распределения ответственности при AI-инциденте — 105
Приложение 45 План независимой методической консультации — 106
Приложение 46 Программа независимой оценки безопасности архива — 106
Приложение 47 Метод проверки институциональных заявлений о сотрудничестве — 107
Приложение 48 Инструмент контроля межъязыковой правовой терминологии — 108
Приложение 49 Метод аудита собственных ошибок AI Observatory — 109
Приложение 50 План устойчивости при прекращении финансирования — 110
Практический указатель рекомендаций — 110
Источники и указатель оснований — 112



Методологическое предисловие

Исследование основано на официальных документах, доступных по состоянию на 22 октября 2026 года, и не содержит результатов собственного международного полевого обследования, которое ещё не проводилось для описываемой модели. Эмпирические величины внутри условных приложений предназначены исключительно для пояснения методики. Отсутствие сведений о системе или стране не оценивается как отсутствие риска. Особое внимание уделяется различению действующего права, необязательных стандартов и авторских рекомендаций.

Протокол различает наблюдение, интерпретацию и причинное утверждение. Подтверждённое сообщение об ограничении операции позволяет описать событие; связь с ИИ требует отдельного основания; вывод о причине и распространённости требует дополнительного дизайна. Исследователь обязан указать, на каком уровне находится каждый результат. Эта дисциплина определяет выбор источников, структуру карточки, правила расчёта и допустимый язык публичного отчёта.

Научные основания и исследовательский дизайн

Обсерватория рассматривается как наблюдательное исследование с документируемым механизмом обнаружения событий. Выборочная методология задаёт область возможного обобщения; контент-анализ обеспечивает воспроизводимость категорий; теория измерений определяет смысл показателей; институциональное исследование оценивает условия независимости. Право и исследовательская этика устанавливают допустимость сбора и раскрытия. Предлагаемый дизайн не предполагает заранее существующей репрезентативной базы. [M1–M3]

Научное направление	Метод	Проверяемый продукт
Выборочное наблюдение	Описание рамки, каналов и исключений	Схема формирования корпуса
Контент-анализ	Независимое кодирование до обсуждения	Таблица согласия и разбор расхождений
Теория измерений	Определение единицы, числителя и знаменателя	Паспорт каждого показателя
Институциональное исследование	Проверка решений при конфликте интересов	След редакционного решения и пересмотра
Право и этика исследования	Оценка допустимости доступа и публикации	Обоснованный режим данных и защиты лица

Наблюдаемая совокупность и единица учёта

Исследование различает сообщение, событие, источник и затронутое лицо. Одно событие может иметь несколько сообщений и источников; один материал может описывать несколько событий. Правило объединения устанавливается до расчёта показателей. При недостатке данных связь сохраняется как предположение, а спорное объединение проверяется в анализе чувствительности. Это предотвращает искусственное увеличение числа случаев за счёт повторных публикаций.

Наблюдаемая совокупность описывается через предмет, период, территорию, языки и каналы. Если собираются публичные решения и обращения, результат характеризует зарегистрированный корпус при данном способе сбора. Для оценки частоты во всех операциях нужен дополнительный знаменатель и обоснованный механизм обнаружения. Неизвестная вероятность включения не исправляется автоматически увеличением объёма базы.

Исследовательские вопросы и проверяемые ожидания

Первый вопрос пилота — могут ли независимые исследователи по одному правилу одинаково включать и классифицировать новые материалы? Второй — устойчив ли описательный вывод к разумным решениям об источниках, дублях и неопределённых случаях? Третий — меняется ли структура зарегистрированных событий после расширения каналов? Последний вопрос сам по себе не устанавливает изменение реального общественного риска.

Для оценки улучшенной инструкции сравниваются независимые результаты до и после её разработки на разных, сопоставимых наборах. Ожидаемое улучшение — уменьшение существенных разногласий без потери содержательно важных категорий. Простое укрупнение классификатора может повысить согласие, одновременно уничтожив нужное различие. Поэтому оцениваются и воспроизводимость, и пригодность категорий для исследовательского вопроса.

Независимая разметка и проверка источника

Инструкция определяет единицу анализа, допустимые категории, правила работы с недостаточными данными и примеры пограничных случаев. Оценщики сначала работают независимо; затем причины расхождений разбираются с сохранением исходных меток. Работа Krippendorff обосновывает необходимость осмысленной проверки надёжности контент-анализа [M1]. Выбор коэффициента согласия зависит от шкалы, числа оценщиков и пропусков; он не заменяет содержательный разбор.

Проверка источника выполняется отдельно от согласованности. Два оценщика способны одинаково классифицировать недостоверное сообщение. Поэтому карточка содержит основание подтверждения, статус первичного материала, дату и существенное противоречие. Отсутствие независимого подтверждения сохраняется как ограничение даже при полном согласии кодировщиков. При разрешении разногласия фиксируется новое основание, а не только итоговый голос.

Сопоставимость и чувствительность выводов

Для временного или международного сравнения проверяются одинаковый смысл категорий, период регистрации, состав каналов и доступность материалов. Результат пересчитывается на стабильном подмножестве источников; отдельно показываются новые каналы и нерешённые случаи. Рекомендации OECD и JRC по индикаторам используются как основание проверять влияние выбора параметров и не скрывать нестабильность сводной оценки. [M2]

Критерий успешности пилота — воспроизводимая процедура с объяснимыми ограничениями и исполнимым исправлением. Уменьшение реального вреда является отдельным исследовательским исходом: для него нужны наблюдения последствий и обоснованное сравнение. Представленные ниже расчёты иллюстрируют методику; собственные результаты международного полевого исследования не заявляются.

1 Предмет исследования и исследовательский вопрос

Институциональная проблема

Предмет исследования — способность независимой обсерватории формировать сопоставимые и проверяемые данные об общественно значимых применениях ИИ. Проверяемость означает возможность другого исследователя восстановить источник, процедуру отбора и основания классификации. Сопоставимость означает совпадение смысла категорий и условий наблюдения в сравниваемых группах. Наличие единого сайта или общей формы карточки само по себе этих свойств не устанавливает.

Проверяемый вопрос

Основной исследовательский вопрос: какие сочетания источников, правил отбора и независимого кодирования позволяют получать устойчивые описательные выводы при ограниченных ресурсах? Вопрос о снижении общественного вреда ставится отдельно и требует данных о последствиях и обоснованного сравнения. Доклад разрабатывает методику наблюдения и программу её пилотной проверки; он не сообщает о доказанной эффективности уже действующей глобальной сети.

Результат и пределы

Доклад предлагает проект устава, методику отбора кейсов, классификатор источников, систему показателей и реестр рисков. Это модель для последующего обсуждения и тестирования, а не заявление о том, что такие органы, процессы или официальные полномочия уже созданы у ARGA. Инструменты ориентированы на трансграничные дела и сектора, где неправомерная автоматизация способна повлиять на свободу передвижения, доступ к финансовым услугам, имущество и процессуальную защиту. Каждое предложение проверяется по трём критериям: публичная полезность, воспроизводимость и стоимость независимого контроля.

Регистрация исследовательского вопроса и критериев вывода

Регистрационная записка фиксирует исследовательский вопрос, вид анализа, источники и критерии достаточности. Для причинного исследования формулируются гипотезы и проверяемые альтернативы; для описательной работы — правила наблюдения и классификации; для правового анализа — юрисдикция, редакция норм и способ толкования. Не каждому документальному обзору нужна причинная гипотеза. В деле об автоматизированной банковской блокировке факт ограничения сам по себе не устанавливает применение конкретной модели: при отсутствии таких сведений вывод может относиться только к подтверждённой процедуре.

Для проверки результативности Observatory предлагается вести журнал исследовательских обещаний: что предполагалось установить, что реально удалось проверить и какие источники остались недоступны. Он защищает от последующего изменения формулировок цели под получившийся результат. Все выводы для внешнего использования проходят тест адресата: кто способен действовать по рекомендации и какие доказательства ему объективно нужны. Если адресату требуется официальный протокол или доступ к нераскрываемым

логам, публикация должна объяснять пределы самого исследования, а не изображать уже выполненную инспекцию.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Наблюдаемый результат	Проверяемая публикация с приложенной методикой
Не является результатом	Необоснованный рейтинг организаций без данных
Институциональный предел	Отсутствие властных полномочий и автоматического доступа к закрытым системам

Ключевые источники для сопоставления: [1], [2], [7], [20].

Практический разбор 20.01

Рекомендация: Задавать исследовательский вопрос до сбора

Условный пример: Команда собирает любые новости об ИИ и хочет оценить распространённость вреда.

Разбор: Новостной корпус не образует автоматически репрезентативную выборку.

Практические действия: Куратор ограничивает предмет, период, языки и измеряемый исход; цель фиксирует до анализа.

Критерий результата: Вывод отвечает заданному вопросу и не выходит за наблюдаемую совокупность.

2 Определение независимого AI Observatory

Три функции

Под Observatory понимается постоянно действующая аналитическая структура, собирающая и проверяющая информацию об общественно значимых последствиях ИИ. Первая функция — документирование: фиксация объектов, событий и доказательств с происхождением и датой. Вторая — сопоставление: расчёт метрик при одинаковых определениях, периодах и ограничениях. Третья — публичное объяснение: отчёты, методические записки, реестры открытых данных и корректировки ранее опубликованных выводов. Для качественного контроля недостаточно вести сайт с новостями: новость становится наблюдением только после классификации, устранения дублей, проверки связи с ИИ и описания неопределённости.

Отграничение от других институтов

Научная комиссия может оценивать состояния знания, государственный надзорный орган — требовать исполнения правовой нормы, орган сертификации — удостоверять установленное соответствие, судебный эксперт — давать заключение по поручению в конкретном процессе. Observatory вправе взаимодействовать со всеми ними, но не присваивает их статус и не приписывает своим выводам юридическую силу, которой они не имеют. Особенно важно различать слова «мониторинг», «аудит», «оценка соответствия» и «расследование»: от них зависят полномочия, ответственность и ожидания адресата.

Рабочая единица наблюдения

Базовой единицей предлагается карточка «система — контекст применения — событие — доказательства». Один и тот же продукт может использоваться с разными данными, порогами и мерами человеческого контроля, а значит образовывать разные объекты оценки. Карточка должна содержать дату среза, географию, владельца решения, версию системы, наблюдаемые последствия, уровень подтверждения и статус ответа затронутой организации. Международные отчёты агрегируют карточки, не стирая их неодинаковую доказательственную силу.

Статус выводов и запрет ложной атрибуции полномочий

Классификация продукции Observatory допускает три уровня утверждения. На уровне наблюдения фиксируется доступный факт: выпущен договор, подана жалоба, вынесен акт, выявлено конкретное несоответствие в открытой документации. На аналитическом уровне делается обоснованный вывод о риске или закономерности, который не обязан иметь силу правового решения. На рекомендательном уровне предлагается изменить практику или исследовательский протокол. Ошибка начинается, когда третий уровень преподносят как обязательное распоряжение или первый как окончательную оценку вины.

Публикационная карточка поэтому должна начинаться с одной фразы о статусе утверждения. Для внешнего читателя особенно опасна лексика «сертифицирован», «расследовано международным органом» или «официально проверено», если соответствующая процедура и полномочие отсутствовали. Предлагается правило двух подписей для текста, содержащего упоминание конкретного государства, банка или международной структуры: методический редактор проверяет основание вывода, юридический редактор — возможное впечатление о статусе автора и его полномочиях. Это недорогая мера, предотвращающая значительную часть репутационных и правовых рисков.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Документирование	Карточка с происхождением данных
Сравнение	Единое определение числителя и знаменателя
Коммуникация	Публичная версия с оговорками
Статус	Частная исследовательская организация

Ключевые источники для сопоставления: [1], [5], [6], [20].

Практический разбор 20.02

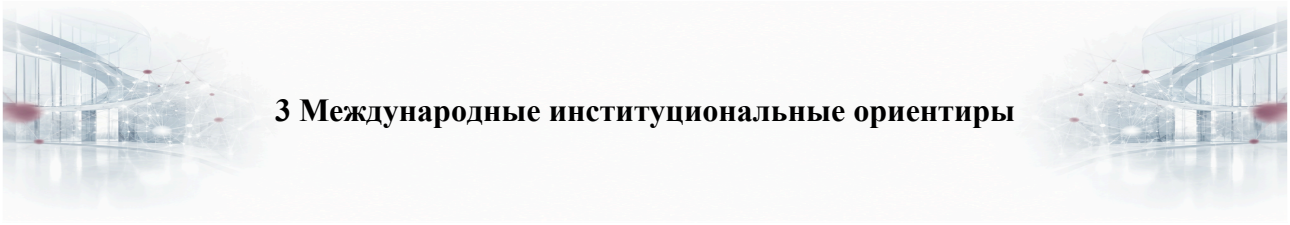
Рекомендация: Разделять наблюдение и властное решение

Условный пример: Публикацию обсерватории предлагают назвать официальным установлением нарушения.

Разбор: Исследовательский статус не создаёт полномочий суда или регулятора.

Практические действия: Редактор описывает наблюдения и рекомендации; юридические выводы ограничивает компетенцией и материалами.

Критерий результата: Публичный текст точно передаёт статус исследования.



3 Международные институциональные ориентиры

ООН

Резолюция Генеральной Ассамблеи ООН 79/325 создала Независимую международную научную группу по ИИ и Глобальный диалог по управлению ИИ. Научная группа состоит из экспертов, действующих в личном качестве, и готовит научные оценки; Диалог обеспечивает многостороннее обсуждение. Эти механизмы не являются всемирным регулятором. Для независимого Observatory полезны их принципы международной репрезентативности, прозрачного мандата и отделения доказательных оценок от политических решений. Взаимодействие возможно только по предусмотренным каналам и не означает институционального членства или признания конкретной организации. [1], [2]

UNESCO и OECD

Глобальная обсерватория этики и управления ИИ UNESCO содержит методологические инструменты и страновые профили; RAM предназначена для оценки готовности государственных систем, EIA — для оценки отдельной AI-системы. OECD.AI объединяет политические инструменты и мониторинг инцидентов/угроз. Это разные масштабы измерения: зрелость государства, риск продукта и документирование события нельзя сводить в один индекс. Предлагаемая модель использует эти подходы как независимые методические ориентиры и сохраняет прямые ссылки на первоисточники. [3]–[5], [7]

Совет Европы и национальные органы

HUDERIA помогает оценивать риски для прав человека, демократии и верховенства права и не является самостоятельной обязательной нормой. Рамочная конвенция Совета Европы CETS 225 существует как договорный инструмент; её обязательность для конкретной страны определяется условиями вступления в силу и участием этой страны, а не фактом публикации конвенции. Национальные регуляторы, суды и омбудсмены остаются адресатами юридических процедур. Observatory может выявить проблему и передать обоснованную информацию, но окончательный вывод о нарушении делает компетентный орган. [9], [10]

Карта функций международных институтов

Сопоставление международных механизмов требует карты функций, а не таблицы логотипов. У научной группы ООН исследовательская функция, у многостороннего Диалога — дискуссионная, у UNESCO — нормативно-методическая и страновая поддержка, у OECD.AI — сравнение политики и событий, у государственных регуляторов — применение национального права. Для каждого возможного контакта следует описать ожидаемый тип вклада Observatory и официальный канал. Например, предоставление агрегированных данных может быть предметом исследовательского обмена, а письменный комментарий в консультации — формой участия без права использовать название учреждения как знак институционального одобрения.

Доказательная база о существовании партнёрства хранится отдельно от реестра дел: официальная программа мероприятия, подтверждение получения комментария, договор или опубликованный совместный материал. Слово «сотрудничество» допускается только с указанием фактического действия. Если наблюдатель по собственной инициативе выпустил обзор на основе методики UNESCO, корректное описание — «использованы принципы и инструменты», а не «исследование подготовлено совместно с UNESCO».

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Группа ООН	Независимая научная оценка
UNESCO RAM	Страновая институциональная готовность
OECD AIM	События, инциденты и опасности
HUDERIA	Контекстно-зависимая оценка прав

Ключевые источники для сопоставления: [1], [2], [3], [4], [5], [7], [9], [10], [21], [22].

Практический разбор 20.03

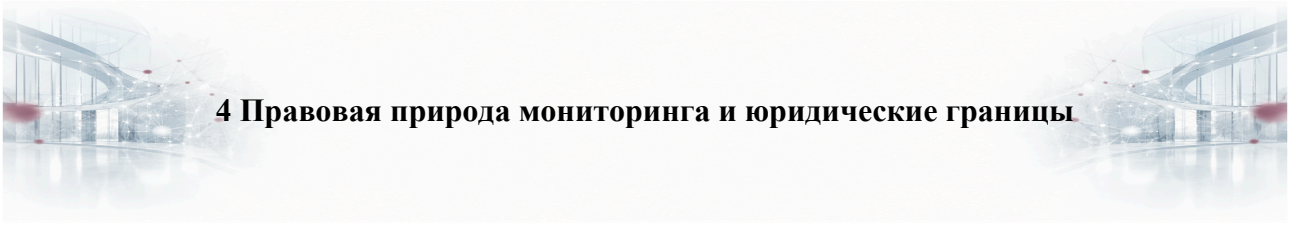
Рекомендация: Переносить институциональный опыт с учётом ресурсов

Условный пример: Небольшая команда хочет скопировать все функции крупного международного портала.

Разбор: Заимствование структуры не обеспечивает финансирования и качества.

Практические действия: Руководитель выбирает ограниченную функцию, доступные источники и необходимую рецензию.

Критерий результата: Пилот выполним имеющимся составом без неподтверждённых заявлений о масштабе.



4 Правовая природа мониторинга и юридические границы

Применимое право

Право на проведение общественного исследования не устраняет требований к защите данных, интеллектуальной собственности, конфиденциальности, диффамации и безопасности. В ЕС анализ персональных данных требует применимого правового основания и соблюдения принципов GDPR. Исследовательская цель не является универсальным освобождением от этих обязанностей. Использование информационных запросов, судебных актов и открытых реестров должно учитывать ограничения вторичного использования и возможность ошибочной идентификации лица. В иных юрисдикциях правовые основания проверяются отдельно; перенос европейского подхода без оговорки создаёт ложную уверенность. [13]

Статус выводов

Вывод «по доступным материалам обнаружен признак алгоритмического риска» отличается от утверждения «организация нарушила закон». Последнее предполагает надлежащий правовой анализ и, в спорных ситуациях, компетентную процедуру. Публикации Observatory должны содержать стандартные формулы степени уверенности: факт документирован; обстоятельство вероятно; гипотеза требует подтверждения; сведения противоречивы; данных недостаточно. Присвоение статуса должно зависеть от доказательств, а не от общественного резонанса. Отдельно указывается, имеет ли оценка характер юридического мнения, научного анализа или технического теста.

Связь с AI Act

EU AI Act устанавливает обязанности для определённых субъектов и систем, а независимый наблюдатель не становится надзорным органом на основании публикации этого акта. После изменений 2026 года сроки применения различных блоков норм различаются; для систем высокого риска предусмотрены отдельные переходные этапы. Поэтому чек-лист Observatory фиксирует применимую дату, категорию системы и роль участника цепочки. Запрещено использовать одну календарную дату как доказательство нарушения любой обязанности AI Act. В сомнительных случаях публикуется карта применимости с ссылкой на актуальную редакцию EUR-Lex.

Правовая проверка перед сбором информации

Для каждой публикации, затрагивающей потенциальное нарушение права, предлагается краткая четырёхчастная юридическая карта: (1) установленная фактическая ситуация; (2) конкретная норма с датой и кругом обязанных лиц; (3) вопрос о применимости этой нормы к установленным фактам; (4) оставшиеся ограничения и компетентный адресат. Если любой шаг не подтверждён, редакционный вывод не должен квалифицировать поведение как нарушение. Это особенно важно для быстро меняющегося AI Act, где разные классы систем и субъекты имеют различные этапы применения требований.

В международном наблюдении необходимо проверять также правила трансграничной передачи информации и режим профессиональной тайны. Даже вполне общественно значимый материал нельзя без анализа публиковать вместе с паспортными данными заявителя или конфиденциальными банковскими документами. Для малой команды полезен отдельный журнал «основание обработки — цель — срок — получатели», который утверждается ещё до открытия онлайн-формы жалоб. Наличие информативной правовой сноски не исправляет незаконность первоначального сбора.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Личный набор данных	Проверка основания и минимизации
Обвинение в нарушении	Юридический анализ и процедура ответа
AI Act	Категория системы, роль и дата применения
Международная публикация	Оценка правовых ограничений каждой юрисдикции

Ключевые источники для сопоставления: [11], [12], [13], [23].

Практический разбор 20.04

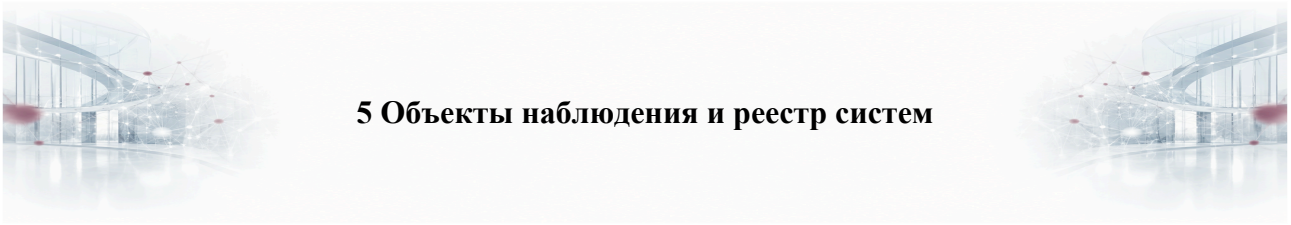
Рекомендация: Проверять правовое основание мониторинга

Условный пример: В публичный реестр планируют включать чувствительные документы заявителей.

Разбор: Получение документа не означает право раскрыть его всем.

Практические действия: Ответственный определяет основания обработки, состав открытых сведений и режим закрытого хранения.

Критерий результата: Публичная карточка содержит только обоснованно раскрываемые данные.



5 Объекты наблюдения и реестр систем

Что включать

Объектом мониторинга становится не «ИИ вообще», а развернутое применение с социально значимой функцией. Приоритет получают системы, которые влияют на доступ к правосудию, миграционный статус, банковские отношения, социальные выплаты, трудоустройство, образование, здоровье и публичную безопасность. Реестр может также содержать инфраструктурные модели общего назначения, но их оценка отличается от оценки конкретного решения. Для каждой записи нужны функциональное назначение, владелец внедрения, тип модели, статус эксплуатации, степень автономии, нормативная категория и географический охват.

Границы инвентаризации

Наличие системы в реестре не означает незаконности и не должно превращаться в публичный «чёрный список». Источниками подтверждения могут быть официальные сообщения, закупочная документация, отчёты об оценке воздействия, судебные акты и технические публикации. Пресс-релиз компании — свидетельство заявления компании, но не независимая проверка реального использования. Когда применимость ИИ не подтверждена, карточка маркируется как предварительная и не включается в статистику внедрённых систем.

Обновление

За жизненный цикл система может изменить алгоритм, провайдера, параметры, категорию пользователей, объём данных и юридическую роль организации. Реестр должен поддерживать версии с датой вступления изменений в силу и историей источников. В противном случае два отчёта за разные годы будут сравнивать разные технические продукты под одним торговым названием. Система архивации обеспечивает обратимость публикаций: пользователь может восстановить, на каком основании карточка имела конкретный статус в прошлом квартале.

Единица учёта и версия наблюдаемой системы

Чтобы исключить произвольное расширение реестра, паспорт AI-системы проходит процедуру входного контроля из трёх подтверждений. Требуется указать фактическую функцию системы; источник её эксплуатации; связь с отдельным контекстом и субъектом. Необязательна публичная информация о коде модели, но тогда уровень технической детализации честно помечается как неизвестный. Для систем с одним брендом и несколькими развертываниями создаются родительская запись продукта и дочерние карточки использования. Иначе изменение решения одной компании ошибочно приписывается всем организациям, купившим её программное обеспечение.

У каждой карточки есть две независимые даты: «наблюдаемый период» и «момент последней проверки». Программа может прекратить работу за месяцы до публикации исследования; оставлять статус «используется» было бы небрежностью. История прекращённых и заменённых систем сохраняется, поскольку именно она позволяет видеть, как учреждения реагируют на инциденты. При обобщении указывается, сколько систем имели подтверждённый эксплуатационный статус на выбранную дату.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Идентификатор	Стабильный ID плюс история версий
Сфера	Назначение и затронутая группа
Применимость	Реальная эксплуатация, не реклама
Время	Дата данных, дата проверки и дата публикации

Ключевые источники для сопоставления: [5], [6], [11], [15].

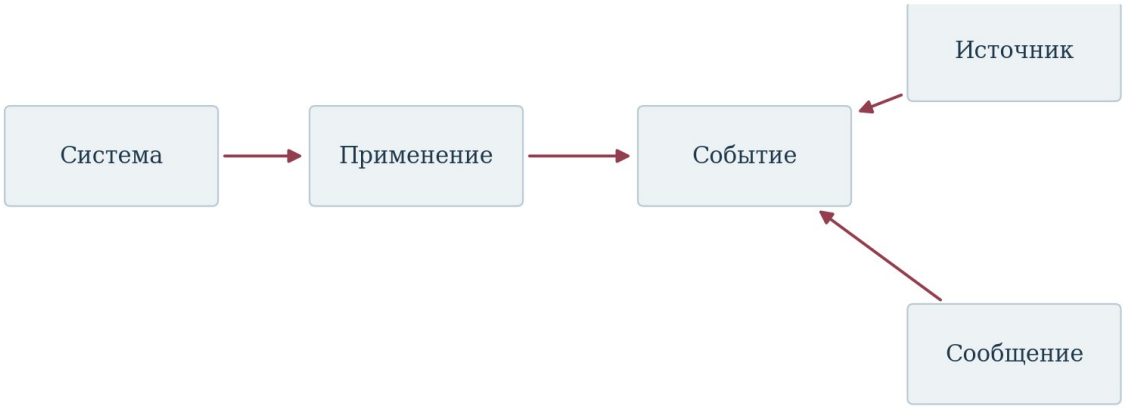


Рисунок 20.1. Связи между объектами наблюдения. Модель раздела 5. Сообщение, событие, источник и затронутое лицо учитываются раздельно; число публикаций не равно числу независимых подтверждений.

Практический разбор 20.05

Рекомендация: Разделять систему и её применение

Условный пример: Одна модель используется банком и миграционным посредником.

Разбор: События относятся к разным процессам и субъектам.

Практические действия: Аналитик создаёт отдельные записи модели, развёртывания и события, связывает их идентификаторами.

Критерий результата: Событие связано с фактическим оператором и назначением.

6 Организационная архитектура

Разделение ролей

Для Observatory предлагаются четыре функционально разделённых звена. Научно-методический совет утверждает классификаторы и проверяет независимость методик. Аналитический секретариат осуществляет сбор и кодирование данных. Комитет по этике и защите данных рассматривает чувствительные кейсы, безопасность и баланс прав. Редакционная комиссия принимает решение о публикации после проверки качества и процедуры ответа. Представители одного проекта не должны одновременно назначать экспертов, вести сбор спорных данных и единолично санкционировать окончательные выводы.

Малый запуск

Полная архитектура не обязательно требует значительного штата на первом этапе. При ограниченных ресурсах разделение ролей достигается письменными полномочиями, независимой внешней рецензией и журналом решений. Например, аналитик создаёт карточку, второй эксперт проводит blind review, юрист проверяет риск раскрытия данных, а редактор утверждает текст. Даже в маленькой команде недопустима ситуация, когда лицо, получающее оплату от проверяемой организации, самостоятельно утверждает оценку этой организации.

Ответственность

Каждое решение должно иметь владельца и срок исполнения: дата регистрации входящего сообщения, дата назначения рецензента, дата уведомления затронутого лица, дата закрытия проверки. Для существенных конфликтов предусматривается внешний независимый арбитр процедуры. Полномочия совета не включают переписывание научных выводов по репутационным мотивам; его задача — обеспечить корректную методику и этику. Возражение совета документируется и при необходимости раскрывается в форме несогласия с методикой.

Распределение полномочий и право остановить публикацию

Распределение полномочий следует закрепить в RACI-подобной матрице, где для ключевых действий указаны исполнитель, ответственный за утверждение, консультируемый специалист и получатель уведомления. В регистрации обращения нужен один ответственный; в юридической оценке — профильная компетенция; в публикации риска для конкретной организации — независимый рецензент. Если сотрудник исполняет две роли, журнал показывает это совмещение и компенсирующий контроль. Признание ограниченных кадровых ресурсов не служит основанием отменить проверку конфликта.

Полномочия институциональных органов должны включать право остановить публикацию по документированному основанию безопасности или грубой фактической ошибке. Но при этом совет не получает право заменить научное заключение «более удобным». Внутренний регламент должен различать остановку ради устранения методической ошибки и попытку давления на автора. Протокол заседания фиксирует основания, срок повторной проверки и ответственного за возвращение материала в работу.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Совет	Методология и независимость
Секретариат	Сбор и хранение данных
Комитет этики	Риски для участников исследования
Редакция	Право на публикацию и исправление

Ключевые источники для сопоставления: [1], [3], [14], [16].

Практический разбор 20.06

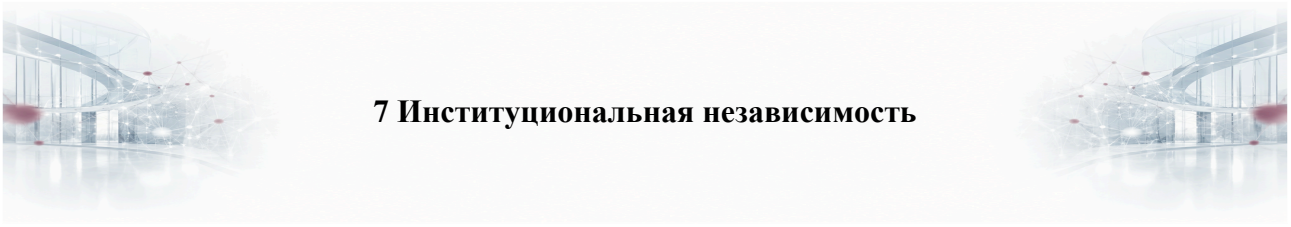
Рекомендация: Назначать разные роли проверки и публикации

Условный пример: Один сотрудник принимает жалобу, проверяет её и утверждает публикацию.

Разбор: Ошибка первого этапа не встречает независимого контроля.

Практические действия: Куратор разделяет подготовку и утверждение; для малой команды привлекает внешнего рецензента.

Критерий результата: Существенная публикация имеет вторую документированную проверку.



7 Институциональная независимость

Независимость как проверяемая система

Независимость организации нельзя установить одной фразой в уставе. Необходимы ограничения на финансирование конкретных выводов, раскрытие конфликтов интересов и правила отвода. Независимость рассматривается в трёх измерениях: финансовом (нет возможности купить благоприятное заключение), кадровом (эксперты не подчиняются проверяемому субъекту) и редакционном (договор не допускает цензуры результатов). Отдельно оценивается независимость от политических групп и органов, заинтересованных в ограничении публичного доступа к сведениям.

Риск невидимого влияния

Донор может формально не требовать изменения текста, но ограничивать темы, сроки или бюджеты. Поэтому публикуются не только денежные поступления, но и существенные in-kind услуги, оплаченный доступ к данным и условия закрытых исследований. При конфликте интересов конкретный эксперт отводится, методика передаётся второму рецензенту, а факт отвода фиксируется в закрытом реестре. Для публичного отчёта допустим агрегированный характер раскрытия, когда персональные данные сотрудника необходимо защитить.

Проверка независимости

Раз в год внешний рецензент оценивает структуру финансирования, количество отводов, случаи давления на публикации и соблюдение процедуры рассмотрения возражений. Желателен специальный показатель: доля проектов, чьи договоры прямо закрепляют независимость редакционного решения. Но высокий процент сам по себе не доказывает отсутствия давления. Итог формулируется как оценка процедур, а не как абсолютная гарантия непредвзятости.

Особая проверка необходима, если исследовательский Observatoire разделяет бренд, руководство, помещения, финансирование или внешних контрагентов со структурой, оказывающей платные правовые, консультационные либо представительские услуги. Общая принадлежность не делает независимые выводы невозможными, но требует раскрытия связи и доказуемого разграничения ролей. Нельзя ограничиваться формальной фразой «редакция независима»: должны существовать письменные ограничения доступа к досье клиентов, отдельные права утверждения публикаций и отвод лиц, заинтересованных в исходе рассматриваемого вопроса.

Рекомендуемый организационный барьер: журнал связанных лиц и оплаченных работ; запрет на финансирование конкретного вывода заказчиком, которому этот вывод выгоден; обязательная внешняя рецензия спорной публикации при пересечении с клиентской работой; публичная политика исправлений, одинаково доступная критикам и сторонникам. Заказчик или донор может указать фактическую ошибку и получить мотивированный ответ, но не получает права вето по репутационным основаниям. При невозможности обеспечить такие меры публикация о соответствующем лице переносится либо выполняется независимой стороной.

Статус независимости описывается фактически: независимость методической рецензии, редакционного решения, бюджета и управления проверяются по отдельности. Наличие нескольких подписей внутри одной административной иерархии не равно независимой внешней проверке. Ежегодный отчёт должен содержать долю отводов, число заявленных конфликтов, долю исследований, связанных с собственными платными услугами, и случаи отклонённого редакционного давления без разглашения персональных сведений.

Ситуация	Минимальное правило независимости
Клиентский кейс	Первичная проверка конфликта, отвод связанных лиц и независимая рецензия до публикации
Донор или подрядчик	Раскрытие существенных условий влияния; отсутствие права одобрять итоговый вывод
Собственная ошибка ARGA	Применение публичной процедуры исправления вне зависимости от репутационного неудобства
Невозможно разорвать конфликт	Документированный отказ от оценки либо передача её независимому исследователю

Стресс-тест реальной редакционной независимости

Проверка независимости проводится через набор контрольных сценариев. Первый: донор просит не включать неудобную страну в обзор. Второй: бывший разработчик предлагает рецензировать собственную систему. Третий: руководитель проекта требует ускорить публикацию до истечения срока ответа проверяемой стороны. Четвёртый: известное исследовательское учреждение просит поставить его логотип до согласования содержания. Во всех случаях правило должно быть ясным заранее, а не формироваться в зависимости от общественной значимости участника.

Предлагается регистрировать существенное вмешательство как отдельное событие институционального риска. Описываются источник, предмет, попытка влияния, принятый барьер и окончательное решение. В годовой публичной версии данные можно агрегировать, не раскрывая конфиденциальную переписку. Сам факт внешнего давления не гарантирует научной безошибочности Observatory, но способность системно ограничивать такое давление — необходимое условие доверия.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Финансовый барьер	Запрет договорного контроля над выводами
Кадровый барьер	Отвод и второе независимое мнение
Редакционный барьер	Защищённое право публиковать методические выводы
Проверка	Ежегодная внешняя оценка процессов

Ключевые источники для сопоставления: [1], [3], [14], [16].

Практический разбор 20.07

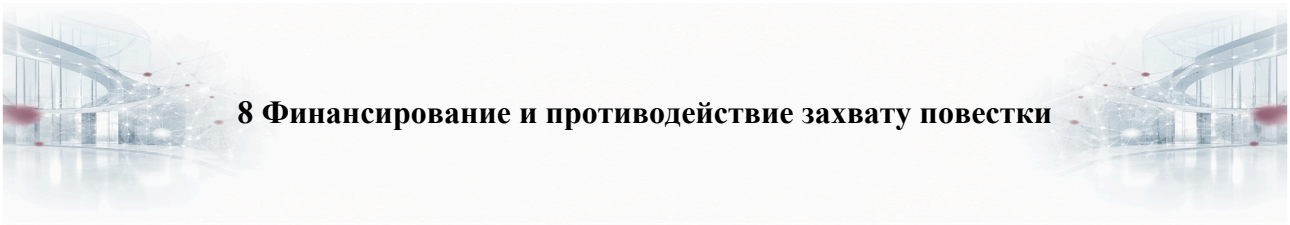
Рекомендация: Защищать право на неудобный вывод

Условный пример: Финансирующий партнёр просит исключить отрицательный результат о его продукте.

Разбор: Редакционное вмешательство ставит независимость под сомнение.

Практические действия: Руководитель применяет заранее согласованные правила независимости и фиксирует запрос.

Критерий результата: Решение о публикации принято по доказательствам и утверждённой процедуре.



8 Финансирование и противодействие захвату повестки

Смешанная модель

Предлагается сочетать базовое институциональное финансирование, целевые гранты, платные методические исследования и пожертвования без права вмешательства в результаты. Базовая часть финансирует реестр, безопасность и публикацию неудобных результатов, которые редко привлекают коммерческий заказ. Целевые гранты допускаются только по заранее определённой тематике и при публично раскрытом характере средств. Платная экспертиза отделяется от общественного мониторинга и не должна превращать свободный от оценки продукт в «одобренный» продукт.

Ограничения доноров

Для предотвращения зависимости предлагается утверждать внутренние пороги концентрации, но не выдавать их за международные правовые требования. Например, ориентир в 20% доходов от одного источника может использоваться как сигнал к повышенной проверке, а не как универсальная граница независимости. Более значимы договорные права донора, структура собственности и способность прекратить финансирование в критический момент. При росте концентрации совет оформляет план диверсификации и раскрывает риск.

Внутренний ориентир концентрации финансирования нельзя превращать в универсальную оценку независимости: доля меньше 20% не исключает редакционного влияния, а доля выше 20% не доказывает зависимость. В методике необходимо одновременно проверять условия договора, право досрочного прекращения гранта, состав органа управления и возможность распределения затрат между независимыми проектами. Условный порог должен регулярно пересматриваться на фактических данных, при этом сам процент и способ его расчёта раскрываются как внутреннее управленческое решение.

Разделение услуг

Организация, одновременно консультирующая разработчика и публикующая оценку того же проекта, несёт повышенный конфликт. Допустимая модель требует отделения команд, запрета использования непубличных данных клиента без правового основания, внешней

проверки чувствительного отчёта и прозрачного обозначения отношения. Не допускается «сертификат безопасности» по итогам заказного обзора, если проверка не соответствовала процессу сертификации и применимым стандартам. Клиенту можно продавать исследовательскую услугу, но не независимость выводов.

Для исследований, затрагивающих лиц, интересы которых ARGA ранее или одновременно представляет, приоритетным вариантом является агрегированное исследование без идентификации человека либо выбор независимого исследования, не использующего информацию клиентского поручения. Наличие подписанного договора о юридической помощи само по себе не разрешает перенос данных в публичный исследовательский реестр. Правомерность вторичной обработки, конфиденциальность и возможный конфликт интересов оцениваются отдельно по каждому набору данных.

Проверка зависимостей от доноров и заказчиков

Для финансовой устойчивости рекомендуется считать не только долю крупнейших доноров, но и коэффициент замещаемости критических функций. Если единственный партнёр предоставляет хранилище, перевод или доступ к реестрам, организация может оказаться зависимой даже при небольшой доле денежных поступлений. Такая зависимость описывается в реестре in-kind поддержки и оценивается по сроку, необходимому для перехода на альтернативного провайдера. Раздельный бюджет исследований, информационной безопасности и юридических консультаций помогает сделать эту угрозу видимой.

Практический тест контракта содержит прямой вопрос: допускает ли соглашение публикацию документированного отрицательного вывода о деятельности самого заказчика? Если нет, результат не следует позиционировать как независимый аудит. Допустимые ограничения, связанные с охраной тайны, уточняются отдельно от цензуры оценочных суждений. При невозможности обеспечить редакционную автономию проект переводится в формат конфиденциальной консультации без внешнего заявления о независимой оценке.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Грант	Публичный предмет и отказ от редакционного контроля
Контракт	Проверка конфликтов до начала работы
Пожертвование	Реестр существенных доноров и условий
Угроза концентрации	Сигнал для совета и диверсификация

Ключевые источники для сопоставления: [3], [14], [16], [17].

Практический разбор 20.08

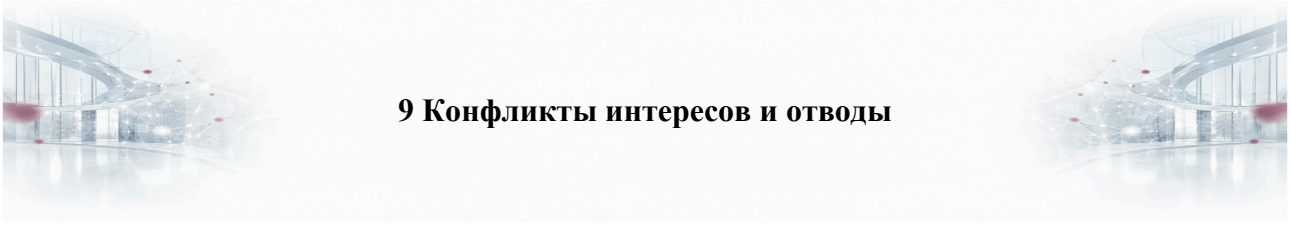
Рекомендация: Раскрывать зависимость финансирования

Условный пример: Большая часть бюджета поступает от одного поставщика наблюдаемой технологии.

Разбор: Зависимость может влиять на выбор тем даже без прямой цензуры.

Практические действия: Совет раскрывает источник, оценивает конфликт и устанавливает независимое утверждение соответствующих исследований.

Критерий результата: Финансирование и меры управления зависимостью доступны для проверки.



9 Конфликты интересов и отводы

Категории

Конфликт интересов может быть финансовым, профессиональным, личным, политико-институциональным и информационным. Эксперт, ранее разрабатывавший модель, не обязательно лишён всякой компетентности, однако его роль требует раскрытия и ограничения при оценке собственных решений. Конфликтом может быть не только прямой контракт, но и оплаченная консультация, спор с проверяемой организацией, владение долей конкурента или обязанность сохранять коммерческую тайну. Реестр должен содержать период возникновения, тип интереса, затронутый проект и принятое решение.

Управление

До назначения эксперт подаёт декларацию. Назначающий сотрудник сравнивает её с реестром заказчиков, анонимизированными сведениями о пожертвованиях и предыдущими проектами. По результату выбирается отсутствие ограничений, участие с раскрытием, ограниченная роль либо полный отвод. При наличии тяжёлого конфликта публикацию ведёт независимый специалист, а первичный эксперт может отвечать только на технические вопросы по документированному запросу. Решение об исключении не должно зависеть от того, выгоден ли вывод организации.

Последующий конфликт

После публикации возможна новая информация о связанных лицах или скрытом финансировании. В таком случае проводится ретроспективный аудит, оценивается влияние на результат и публикуется объяснение возможной корректировки. При невозможности подтвердить независимость ключевого рецензента соответствующее утверждение снимается либо маркируется как требующее повторной проверки. Обстоятельства защищаются от необоснованного раскрытия персональных данных; достаточно показать характер влияния и процесс управления им.

Протокол отвода и документирование конфликта

Конфликт интересов оценивается не по моральному отношению к эксперту, а по механизму возможного влияния на результат. Например, специалист может иметь ценную отраслевую компетенцию, полученную при работе в проверяемой компании. Его участие в объяснении технического контекста допустимо при раскрытии связи, однако решающую оценку собственной прежней работы должен проводить другой рецензент. Аналогично острый публичный спор с объектом исследования способен привести к заведомому негативному отбору фактов.

При годовой оценке конфликтов полезно считать не число деклараций, а долю назначений, прошедших содержательную проверку, и случаи нарушения решений об отводе. Внешний проверяющий получает доступ к обезличенному реестру, чтобы оценить процедуру, не распространяя личные финансовые сведения сотрудников. Если новый конфликт обнаружен после публикации, нельзя просто дописать оговорку: необходимо определить, затронул ли он выбор источников, расчёт и итоговый вывод.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Декларация	До начала каждого нового проекта
Отвод	Документированное решение без конфликта
Ретроспектива	Пересмотр при раскрытии новых отношений
Публичность	Достаточное раскрытие без избыточных данных

Ключевые источники для сопоставления: [1], [14], [16].

Практический разбор 20.09

Рекомендация: Применять отвод по конкретному вопросу

Условный пример: Рецензент консультировал организацию, которая фигурирует в исследовании.

Разбор: Общая декларация интересов не решает вопрос участия.

Практические действия: Координатор назначает независимого рецензента для спорной части и фиксирует отвод.

Критерий результата: Заинтересованный участник не утверждает вывод по своему вопросу.

10 География и сопоставимость юрисдикций

Условия международной сопоставимости

Международное сопоставление начинается с проверки общей единицы наблюдения. В одной стране источники могут описывать обращения, в другой — решения регулятора, в третьей — отдельные эпизоды вреда. Перед сравнением составляется таблица соответствия категорий и правил включения. Случаи, которые невозможно привести к сопоставимому смыслу без потери существенной информации, анализируются отдельно. Юрисдикционные различия не скрываются внутри одинакового названия показателя.

Межъязыковая проверка категорий

Для общего корпуса предлагается межъязыковая проверка: исследователи разных языковых групп независимо классифицируют один набор переведённых материалов, сохраняя оригиналы и спорные термины. Расхождения разбирают по источнику — перевод, правовое понятие, недостаток данных или неоднозначность инструкции. Успешность на таком наборе проверяет согласованность процедуры, но не устраняет различия доступности реальных источников между странами.

Сопоставление через интервалы

Вместо ранжирования стран одним числом предлагается публиковать профиль и интервалы по группам метрик. Там, где различаются правовые основания, оценка остаётся качественной с ясным объяснением несовместимости. Международные агрегаты считаются только для стран, достигших минимальных требований к отчётности, и сопровождаются списком исключённых юрисдикций. Такой подход менее эффектен для заголовков, зато значительно надёжнее при использовании результатов юристами и международными организациями.

Сопоставимость страны, периода и правовой базы

Для сопоставления стран рекомендуется заранее выпустить протокол отбора юрисдикций. Допустимы выбор по региональному балансу, по зрелости публичных реестров или по наличию конкретной автоматизированной функции, но итог отчёта зависит от такого выбора. Если анализ ограничен странами с развитой судебной открытостью, он отражает доступность доказательств, а не обязательно лучшее или худшее состояние AI governance. Каждый международный график должен иметь под ним перечень стран и ключевые причины исключения остальных.

В правовом профиле полезно отдельно кодировать «норма принята», «норма начала применяться», «орган компетентен», «публичное исполнение документировано». Эти четыре признаки могут принимать разные значения одновременно. Время также важно: нормативная ситуация на дату инцидента и на дату исследования может различаться. Observatory должен сохранять обе версии и не оценивать прошлое действие по новой норме без специального правового основания.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Право	Статус акта и дата применимости
Институты	Компетенции и процедуры жалоб
Практика	Судебные и регуляторные подтверждения
Данные	Полнота и ограничения наблюдения

Ключевые источники для сопоставления: [4], [5], [6], [10], [11].

Практический разбор 20.10

Рекомендация: Сопоставлять страны при одинаковом охвате

Условный пример: В одной стране изучают судебные решения, в другой только новости.

Разбор: Различия числа случаев могут отражать способ сбора.

Практические действия: Исследователь формирует сопоставимый слой источников и отдельно показывает остальные материалы.

Критерий результата: Межстрановой вывод сопровождается проверкой наблюдаемости.

11 Стратегия общественного мониторинга

Три канала сбора

Мониторинг строится из трёх независимых каналов: регулярный обзор публичных документов и судебных решений; добровольные обращения и сообщения организаций гражданского общества; целевые исследования систем, выбранных по заранее объявленному рискоориентированному плану. Нельзя подменять третью группу случайными медийными сюжетами, поскольку так возникает эффект сенсационности. Каналы сохраняют собственные показатели охвата и доли непроверенных сообщений. Приоритетность исследований определяется потенциальной тяжестью вреда, числом затрагиваемых лиц и доступностью независимой проверки.

Открытые и ограниченные данные

Режим открытых источников не предполагает автоматического массового копирования личных данных. Observatory хранит краткие выдержки, библиографию, дату доступа и криптографический отпечаток проверяемых материалов, если это оправдано. Доступ к конфиденциальным данным осуществляется только при наличии правовой и договорной основы, описанного срока хранения и распределения доступа. Результаты публикуются после удаления идентификаторов либо в виде агрегатов, если сохранение индивидуальных сведений не требуется для защиты публичного интереса.

Редакционный цикл

Каждый цикл состоит из регистрации наблюдения, предварительной квалификации, проверки относимости к ИИ, поиска независимого подтверждения, оценки вреда, направления запроса о комментарии и решения о публикации. Для систематических обзоров публикуются правила включения и исключения, а не только истории, которые оказались наиболее эффектными. Если наблюдение не удовлетворяет стандарту доказательств, его можно включить в закрытую очередь исследования, но не в раздел подтверждённых случаев.

Предварительная регистрация мониторинговой программы

Практическое правило работы с несколькими каналами состоит в создании общей модели событий при сохранении метаданных источника. Один судебный акт, два газетных сообщения и обращение гражданина могут относиться к одной истории; в агрегате это одна карточка события с несколькими подтверждениями. В то же время серия сходных ошибочных решений одного ведомства может требовать отдельного учёта эпизодов. Выбор единицы устанавливается до вычисления трендов.

Для планирования мониторинга можно составить карту «сектор — источник — частота проверки». Например, судебные документы проверяются в определённом периоде; сообщения заявителей поступают непрерывно; исследовательская выборка определяется заранее. Недостаток ресурса отражается в показатель покрытия, а не скрывается. Без такой карты невозможно отличить изменение числа событий от изменения интенсивности самой работы Observatory.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Регулярный канал	Отчёты и судебные источники
Обращения	Защищённая форма входящих сообщений
Целевая выборка	Заранее фиксируемые критерии отбора
Публикация	Верификация и право на ответ

Ключевые источники для сопоставления: [7], [8], [13], [20].

Практический разбор 20.11

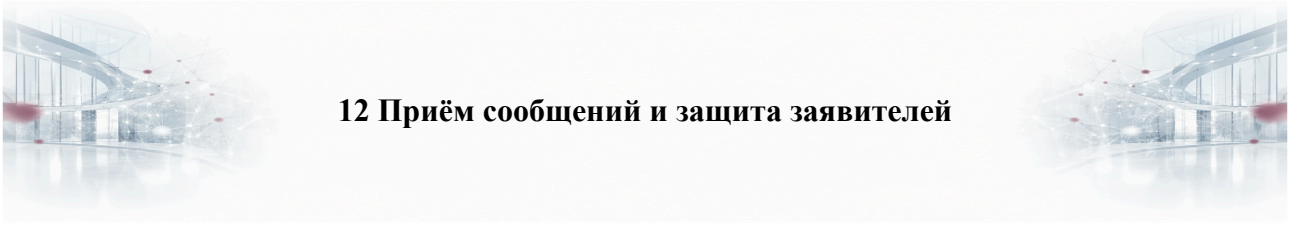
Рекомендация: Фиксировать каналы мониторинга

Условный пример: В середине года добавили два языка и вдвое больше источников.

Разбор: Рост регистрации может объясняться расширением охвата.

Практические действия: Аналитик ведёт журнал изменений и рассчитывает ряд по стабильным каналам отдельно.

Критерий результата: График отмечает смену охвата и не выдаёт её за рост риска.



12 Приём сообщений и защита заявителей

Форма обращения

Форма приёма должна запрашивать минимальный набор сведений: описание решения, контекст применения ИИ, приблизительную дату, вид предполагаемого вреда и доступное подтверждение. Не следует требовать паспорт, адрес и полный массив переписки до оценки необходимости. Пользователь выбирает согласие на обратную связь и отдельное согласие на цитирование; эти действия не заменяют законного основания обработки данных, но помогают управлять ожиданиями. Для сообщений о риске преследования предусматривается анонимный либо псевдонимный канал.

Проверка надёжности

Поступившее сообщение не считается установленным инцидентом. Аналитик отделяет собственный опыт заявителя от предположений о внутренней работе системы, проверяет последовательность дат и наличие альтернативных объяснений. Если сообщение касается действий государственного органа или банка, следует запросить доступные решения, письма, сведения о человеческой проверке и механизме оспаривания. При невозможности верификации наблюдение остаётся конфиденциальным и не влияет на публичную статистику подтверждённых нарушений.

Защита от ущерба

Публикация может причинить дополнительный вред заявителю: раскрыть миграционный статус, медицинские сведения, финансовое положение или участие в процессе. Поэтому проводится оценка угрозы повторной идентификации по сочетанию косвенных признаков. Контактные данные хранятся отдельно от исследовательских карточек, доступ к ним журналируется, а сроки удаления фиксируются заранее. Принцип публичного интереса не означает безусловного права раскрыть всю историю конкретного человека.

Безопасный приём сообщений и границы конфиденциальности

Для сообщений от пострадавших предлагается двухуровневая процедура. На входе принимается минимальный сигнал без требования подтверждать личность. Если событие имеет исследовательский потенциал, ответственное лицо объясняет цели и ограничения, уточняет допустимый контакт и только затем запрашивает дополнительные документы. Заявителю предлагается право не участвовать в публичной истории без автоматической потери возможности получить информацию о доступных официальных процедурах.

Техническое решение должно предотвращать случайное распространение вложений: закрытый канал, отдельное хранение, минимум уведомлений с содержимым дела. При высокой угрозе заявителю допускается передача общей обезличенной информации без персонального контакта. В отчёте нельзя обещать адвокатское представительство, удаление записи или изменение государственного решения лишь на основании регистрации обращения в Observatory.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Минимум данных	Не запрашивать документы по инерции
Статус сообщения	Установлен/вероятен/не подтверждён
Публикация	Отдельная оценка повторной идентификации
Срок хранения	Привязать к цели обработки и процедуре

Ключевые источники для сопоставления: [7], [8], [13], [18].

Практический разбор 20.12

Рекомендация: Разделять приём сообщения и подтверждение

Условный пример: Заявитель прислал эмоциональный рассказ без первичного решения.

Разбор: Рассказ важен для проверки, но не подтверждает все обстоятельства.

Практические действия: Принимающий сохраняет сообщение, уточняет доступные документы и объясняет порядок безопасной передачи.

Критерий результата: Карточка содержит статус проверки и следующий конкретный запрос.



13 Классификация событий: инцидент, опасность, сигнал

Таксономия OECD

Методология OECD различает AI incident, когда развитие или использование ИИ привело к реальному вреду, и AI hazard, когда вероятен вред при определённых условиях. Observatory сохраняет это различие и вводит дополнительный внутренний класс «сигнал», если сведения пока не позволяют установить ни один из статусов. Важно не превращать отсутствие установленных последствий в доказательство безопасности. Один сигнал может перейти в hazard или incident после получения документов, но история этой переклассификации сохраняется.

Связь с технологией

Для статуса AI-события нужно объяснить роль системы: прямое принятие решения, рекомендация, автоматическое ранжирование, генерация текста или косвенное влияние на процесс. Например, ошибочное решение чиновника не становится AI-инцидентом лишь потому, что в учреждении используются цифровые сервисы. Напротив, алгоритмическая рекомендация может быть причинно значимой даже при формальном участии человека. Ключевой вопрос — какие альтернативные решения и фактические действия были доступны оператору.

Объединение дублей

Медийные публикации о той же истории могут создать видимость десятков независимых инцидентов. Для дедупликации используются событие, организация, время, география и первичный документ, однако объединение должно быть осторожным: одинаковые ошибки разных лиц иногда представляют серию отдельных решений. Базовая карточка события связывает дочерние эпизоды и указывает правило подсчёта. Нельзя менять это правило между отчётами без переоценки временного ряда.

Разделение фактического вреда и потенциальной угрозы

При классификации причинно значимым считается не каждое использование автоматизации, а конкретная связь с возникшим или возможным вредом. Аналитик формулирует конкурирующие сценарии: ошибка исходных данных, непригодный алгоритм, неправильная настройка оператора, отсутствие независимого пересмотра и самостоятельное человеческое решение. При отсутствии внутренней документации может быть установлен вред и недоказана роль конкретного AI-компонента; эти статусы фиксируются раздельно.

Для трендов инцидентов и hazards предлагается публиковать матрицу переходов статусов. Если событие на момент поступления было сигналом, затем стало hazard и наконец incident после судебного решения, первоначальные версии не удаляются. Иначе исторический отчёт будет выглядеть так, будто редакция с самого начала обладала окончательной информацией. Пересмотр статуса имеет собственную дату и доказательное основание.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
AI incident	Документированный актуальный вред
AI hazard	Правдоподобная возможность вреда
Сигнал	Недостаточно доказательств
Дедупликация	Стабильное правило события и подэпизодов

Ключевые источники для сопоставления: [7], [8], [19].

Практический разбор 20.13

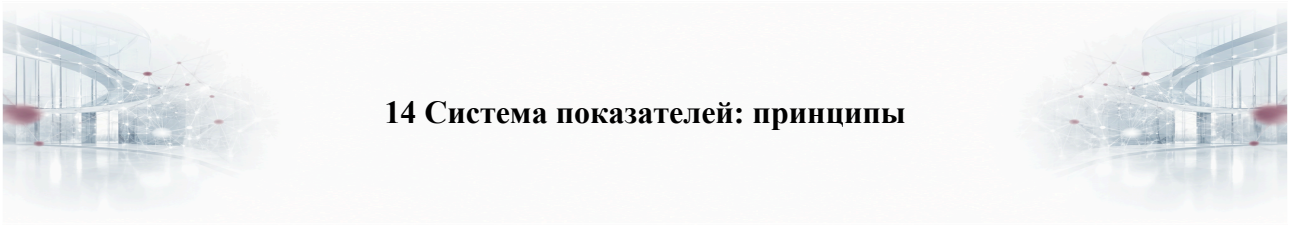
Рекомендация: Различать сигнал опасность и инцидент

Условный пример: В публикации описан возможный ущерб, но нет сведений о его наступлении.

Разбор: Потенциальный сценарий и реализованное событие требуют разных категорий.

Практические действия: Кодировщик применяет определения, сохраняет основание и передаёт спорный случай на рецензию.

Критерий результата: Категория подтверждена наблюдаемыми признаками, неопределённость сохранена.



14 Система показателей: принципы

Качество важнее числа

Каждый показатель получает исследовательское определение: объект, числитель, знаменатель, период, источник и допустимую область вывода. «Число сообщений», «число уникальных событий», «число подтверждённых эпизодов» и «число затронутых лиц» описывают разные величины. Их нельзя складывать или сопоставлять как взаимозаменяемые меры риска. Карточки должны позволять восстановить переход от первоначального сообщения к итоговой единице учёта.

Выбор показателей для исследовательской задачи

Качество показателя оценивается по соответствию исследовательскому вопросу. Для нагрузки редакции полезно число входящих сообщений; для структуры зарегистрированных событий — распределение категорий; для распространённости в популяции нужен обоснованный знаменатель и механизм отбора. Предлагается начинать с раздельных показателей. Составной индекс потребует дополнительного обоснования весов, обработки пропусков и чувствительности, а также объяснения того, какое решение он поддерживает. [M2]

Интерпретация

Если зарегистрированные жалобы выросли с 20 до 40, без числа пользователей и изменений канала обращения нельзя заключить, что риск удвоился. Возможно, увеличились доверие к процедуре и доступность заявлений. Поэтому dashboard обязан показывать показатели полноты выявления и качество источников. При невозможности оценить знаменатель публикуются абсолютные числа наблюдений с пометкой «не показатель распространённости». Подмена такого числа частотой нарушения считается методической ошибкой.

Паспорт показателя и проверка знаменателя

Для каждого KPI формируется карточка с примерами допустимого и недопустимого использования. Например, «доля решений с доступной процедурой обжалования» требует ясного определения: считается наличие правовой нормы, опубликованной инструкции или подтверждённая возможность человека реально получить пересмотр? Эти три показателя несравнимы. Можно публиковать все три, но с разными именами. Подмена их общим термином «уровень защиты» даст ложный вывод.

Отдельный KPI качества Observatory — доля опубликованных карточек, у которых можно открыть первичный документ или получить его описание с обоснованным ограничением доступа. Он полезнее числа пресс-релизов. Для конфиденциальных материалов достаточно результата независимой документальной проверки, но статус такого подтверждения должен быть ясно отличим от полностью публичного источника.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Регистрируемый инцидент	Число подтверждённых карточек за период
Охват	Доля обработанных публичных источников
Процедуры	Доля случаев с документированным ответом
Последствия	Результат пересмотра с указанием недостающих данных

Ключевые источники для сопоставления: [6], [7], [8], [14], [19].

Практический разбор 20.14

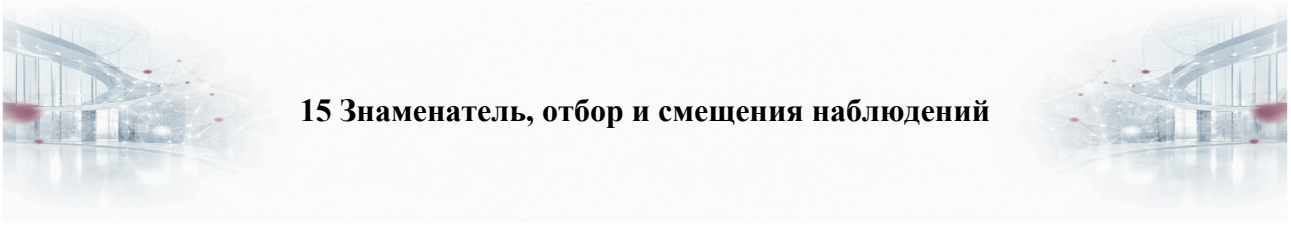
Рекомендация: Определять смысл показателя до расчёта

Условный пример: Панель показывает число пострадавших, считая количество сообщений.

Разбор: Один человек мог отправить несколько сообщений.

Практические действия: Аналитик разделяет сообщение, событие и лицо; вводит правила объединения и пропусков.

Критерий результата: Показатель имеет определение, единицу и воспроизводимый расчёт.



15 Знаменатель, отбор и смещения наблюдений

Проблема знаменателя

Доля или частота рассчитывается только после согласования числителя и знаменателя. Если зарегистрированы 30 подтверждённых эпизодов, но неизвестно число соответствующих решений и полнота обнаружения, допустим вывод о найденных эпизодах. Оценка риска для всех пользователей не следует из этих данных. Даже при известном числе операций необходимо проверить совпадение периода, единицы учёта и охвата; число клиентов за год не заменяет число решений за месяц.

Смещённые источники

Источники образуют разные механизмы отбора. Жалобы зависят от доступности обращения и готовности заявителя раскрыть сведения; новости — от редакционного интереса и языкового охвата; официальные решения — от полномочий и приоритетов органа. Совпадение события в нескольких каналах усиливает его документирование, но не устраняет автоматически смещение отбора. По каждому каналу сохраняются происхождение, правила обнаружения и связь с уже зарегистрированными событиями.

Чувствительность

Анализ чувствительности показывает, насколько вывод зависит от спорных решений исследователя. Отдельно пересчитываются результаты при включении и исключении неопределённых событий, альтернативном объединении дублей и ограничении стабильным набором источников. Сценарии задаются содержательно, а не только произвольным процентом изменения. Если направление вывода меняется, публикуются условия этого изменения и конкретная дополнительная проверка, способная разрешить неопределённость.

Анализ ошибок отбора и невозможности экстраполяции

Сопоставление периодов требует сохранить единицу наблюдения и механизм сбора либо объяснить их изменение. Удвоение зарегистрированных случаев после запуска нового языка может отражать расширение обнаружения. Для анализа строится отдельный ряд по общим источникам и отдельно показывается вклад новых каналов. Такой разбор не восстанавливает неизвестные события прошлого и не доказывает причинный эффект расширения наблюдения.

При неизвестной вероятности включения каждого события большая база не становится репрезентативной автоматически. Коррекция весами требует обоснованной информации о механизме отбора; произвольные веса не устраняют смещение. Допустимым результатом может быть описание структуры зарегистрированного корпуса и анализ механизмов отдельных случаев. Заявление об общей частоте риска откладывается до появления подходящего дизайна и знаменателя.

Изменение числа событий при расширении источников

Условный пример показывает, почему рост зарегистрированных событий не равен росту риска. В первом периоде обнаружены 20 подтверждённых событий по установленному

набору источников. Во втором периоде введён дополнительный языковой канал; общее число событий выросло до 40. После устранения дублей вклад каналов выглядит следующим образом.

Набор источников	Первый период	Второй период
Общие для двух периодов	20	22
Новый языковой канал	Не собирался	18
Всего зарегистрировано	20	40

Общее число увеличилось на 100%, тогда как по общим источникам рост составляет 10%. Второе число тоже не является оценкой изменения реального риска: могли измениться активность публикаций, доступность документов и число соответствующих операций. Корректный вывод описывает увеличение зарегистрированного корпуса и вклад расширения наблюдения. Сравнение реальной частоты потребует сопоставимого знаменателя и исследования полноты обнаружения.

В отчёте сохраняются оба ряда и дата изменения сбора. Новые события не удаляются ради сопоставимости, но аналитический вывод по стабильному набору отделяется от полного объёма регистрации. Если общие источники также существенно изменили практику публикации, даже этот ряд получает дополнительное ограничение. Прозрачность таких изменений важнее видимости непрерывного графика.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Доступный знаменатель	Публиковать частоту с доверительным интервалом
Нет знаменателя	Оставлять абсолютное число наблюдений
Источники неоднородны	Отдельные серии и ясные предупреждения
Слабая устойчивость	Интервал вместо точечного вывода

Ключевые источники для сопоставления: [7], [8], [19], [20].

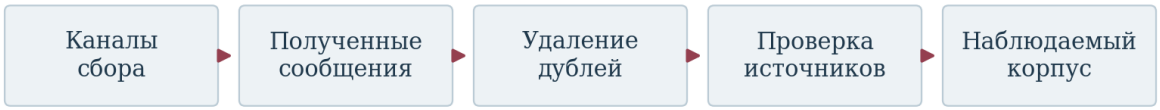


Рисунок 20.2. Отбор корпуса и границы обобщения. Схема раздела 15 без вымышленных количеств. При неизвестной вероятности включения число карточек не определяет распространённость вреда.

Практический разбор 20.15

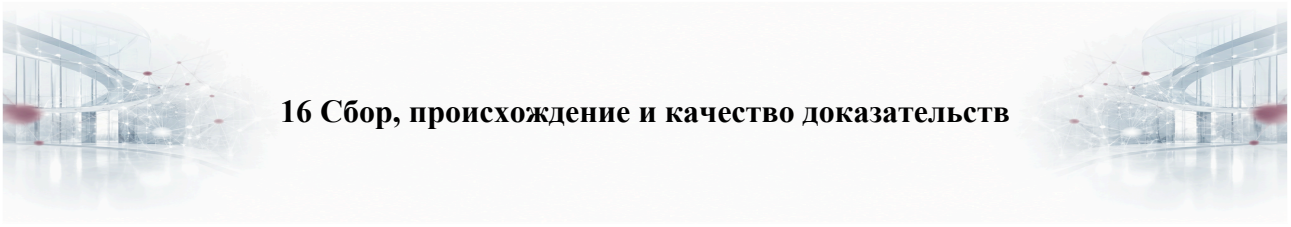
Рекомендация: Указывать знаменатель и механизм отбора

Условный пример: Из десяти обратившихся восемь сообщили о банковской блокировке.

Разбор: Это доля обращений, а не доля всех клиентов с блокировкой.

Практические действия: Статистик подписывает наблюдаемую совокупность и не переносит результат на всю популяцию.

Критерий результата: Внешний текст содержит корректный знаменатель и ограничения отбора.



16 Сбор, происхождение и качество доказательств

Цепочка происхождения

Доказательство включает не только файл, но и сведения о происхождении, времени получения, условиях доступа и преобразованиях. Observatory ведёт журнал поступления материала, присваивает устойчивый идентификатор, сохраняет исходник отдельно от рабочей копии и фиксирует производные операции. Хэширование помогает выявлять изменение электронного объекта, но не доказывает достоверность его содержания или законность получения. Для важных утверждений требуется независимый источник либо проверка непосредственно в официальном реестре.

Стандарты качества

Предлагается четырёхступенчатая шкала подтверждения: первичный официальный документ; прямое документальное свидетельство участника с сопутствующими материалами; независимый вторичный источник с проверяемыми ссылками; непроверенное сообщение. Шкала определяет степень уверенности в конкретном утверждении и не должна превращаться в автоматический рейтинг правдивости источника в целом. Один судебный акт может удостоверять факт судебного решения, но не все утверждения стороны, процитированные в нём.

Воспроизводимость

Исследовательский пакет должен позволять второму специалисту воспроизвести не только вычисление, но и логику классификации. Для закрытых данных сохраняется обезличенный журнал признаков и процедура контролируемого доступа, если это допустимо. В открытом отчёте приводятся источник каждой метрики, выборка, исключения, версия кода и дата проверки. При недостатке данных следует прямо отказаться от сильного вывода, вместо того чтобы маскировать пробел экспертной уверенностью.

Цепочка происхождения и сохранности доказательств

Цепочка доказательств включает как минимум четыре самостоятельных вопроса: откуда получен документ; изменился ли файл; что текст подтверждает; каким правом исследователь может его использовать. Контрольная сумма решает лишь второй вопрос и частично облегчает проверку сохранности. Например, подписанное изображение может быть технически неизменным и содержать фактически ложный текст. Поэтому качество источника оценивается содержательно, а не только программным инструментом.

Цифровые материалы должны иметь воспроизводимую цепочку происхождения: источник и законность получения, момент приёма, исходное имя файла, контрольную сумму, доступы и результаты любых преобразований. Контрольная сумма подтверждает неизменность конкретной копии, но не истинность содержания, полномочия автора или полноту экспортированной переписки. Для скриншотов дополнительно сохраняются URL, дата и контекст страницы, а для AI-генерируемых ответов — точная версия системы, параметры запроса и ограничения повторного запуска.

Для исследований, связанных с судом, рекомендуется готовить пакет с хронологией, нумерацией оригиналов, описанием переводов и журналом маскирования персональных сведений. В научном отчёте указываются пределы доступа: публикуемый отрывок может быть достаточным для проверки одного тезиса, но не всего дела. Если второй исследователь не может воспроизвести классификацию из доступных материалов, результат отмечается как невоспроизводимый извне.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Исходник	Сохранить отдельно от рабочих копий
Метаданные	Время, канал, доступ и обработка
Внешнее подтверждение	Проверить независимо, где возможно
Воспроизводимость	Версии методов и критерии исключения

Ключевые источники для сопоставления: [7], [8], [13], [14].

Практический разбор 20.16

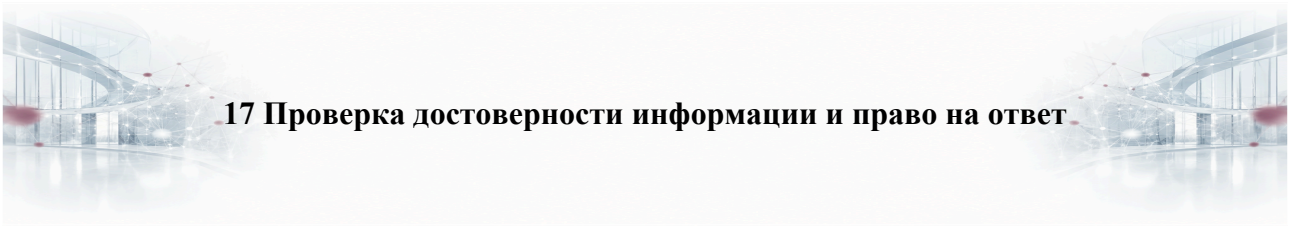
Рекомендация: Прослеживать источник каждого существенного факта

Условный пример: Пять сайтов повторяют один исходный пост.

Разбор: Количество публикаций не означает пять независимых подтверждений.

Практические действия: Исследователь строит цепочку происхождения и выделяет первичный материал.

Критерий результата: Карточка показывает число независимых оснований, а не перепечаток.



17 Проверка достоверности информации и право на ответ

Протокол верификации

Каждое общественно значимое утверждение проходит три этапа: проверка фактического основания, юридико-технический контекст и оценка возможного ущерба от публикации. Редактор не может скрывать факт наличия опровергающих материалов, даже если общая гипотеза выглядит убедительной. Для крупных расследований полезна матрица «тезис — доказательство — альтернативное объяснение — оставшаяся неопределённость». Такая матрица защищает от эффектных, но логически непрочных причинных объяснений.

Ответ затронутой стороны

Организации, в отношении которой готовится существенный негативный вывод, предоставляется разумная возможность ответить на конкретные вопросы. Срок зависит от

сложности запроса и срочности публичного риска; внутренний ориентир устанавливается процедурой, а не объявляется законом. Ответ не даёт право запретить публикацию, но должен быть добросовестно учтён. Молчание не считается признанием обвинений и отражается нейтрально: «ответ не получен на дату закрытия материала».

Работа с опровержением

После публикации новый судебный акт или технический протокол может изменить оценку. Исправление сопровождается датой, описанием изменённого утверждения и сохранением доступной истории, если это правомерно. При грубой ошибке, способной повлиять на жизнь человека, корректировка должна быть заметной и оперативной. Институциональная репутация лучше защищается прозрачным признанием неопределённости, чем попыткой незаметно переписать карточку события.

Разрешение противоречий и право на мотивированный ответ

Право на ответ — часть методической точности, а не уступка объекту исследования. Сторона может раскрыть документ, который опровергает связь с алгоритмом, укажет на ошибку в дате или покажет, что принятое решение было предварительным. Запрос должен быть достаточно конкретным, чтобы ответ мог изменить результат. Слишком общий вопрос «прокомментируйте обвинения» не обеспечивает качество. При этом редакция вправе отклонить мнение, если оно не подкреплено источниками.

Внутреннее решение о публикации содержит таблицу всех спорных фактов, ответа стороны и аналитической оценки ответа. Если не получено ни одного подтверждения прямой связи ИИ с вредом, публикация может обсуждать сам процесс, но не утверждать доказанность технического дефекта. Срочная безопасность заявителя иногда ограничивает объём запрашиваемых сведений; такое исключение документируется и пересматривается после устранения риска.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Тезис	Фактическое основание
Возражение	Отдельная запись в рецензии
Молчание	Не считать подтверждением
Корректировка	Версия, дата, причина, уведомление

Ключевые источники для сопоставления: [13], [18], [19].

Практический разбор 20.17

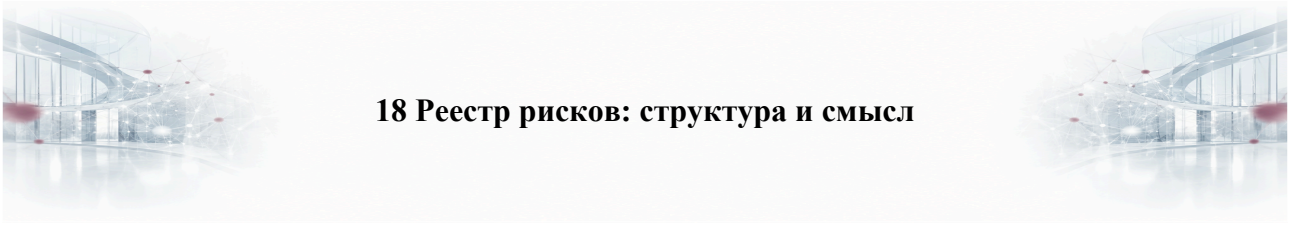
Рекомендация: Предоставлять содержательную возможность ответа

Условный пример: В отчёте готовится вывод о действиях организации без запроса её позиции.

Разбор: Неучтённый документ может изменить квалификацию.

Практические действия: Редактор направляет конкретные вопросы и допустимые сведения, фиксирует ответ либо отсутствие ответа.

Критерий результата: Позиция адресата рассмотрена по существу и отражена в версии публикации.



18 Реестр рисков: структура и смысл

Две единицы учёта

Для устойчивой работы нужны отдельные реестр систем и реестр рисков. Риск относится к сценарию «система — действие — уязвимость — возможный ущерб — затронутая группа», а не только к продукту. Например, ошибочная транслитерация фамилии при санкционном совпадении и нарушение доступа к финансовой услуге представляют разные звенья причинной цепи. В реестре фиксируются источник, контрольная мера, остаточный риск и ответственная сторона. Риск может существовать без подтверждённого инцидента, но это не позволяет выдавать прогноз за факт.

Оценка

Предлагается раздельно оценивать тяжесть возможного вреда, вероятностную уверенность, распространение системы, уязвимость затронутых лиц и обратимость последствий. Сложение условных баллов в единый «индекс опасности» допускается только как внутренняя сортировка и обязательно раскрывает субъективность весов. Важные риски с низкой вероятностью, но необратимым вредом должны иметь собственный порядок эскалации. Методика не заявляет претензии на нормативную юридическую классификацию.

Динамика

После обновления модели, изменения закона или получения новой жалобы карточка риска пересматривается. История включает предыдущую оценку, причину изменения и дату независимого подтверждения. Уменьшение оценённого риска может отражать новую защитную меру, улучшение знаний или изменение группы пользователей; эти случаи принципиально различаются. Механизм обратной связи от организаций и пострадавших позволяет исправлять картину без отмены независимости Observatory.

Разграничение риска, тяжести и уверенности

В реестре рисков степень уверенности нельзя механически смешивать с тяжестью. Одно нарушение, способное привести к необратимому лишению свободы, может требовать срочной проверки при ограниченных данных; частая ошибочная сортировка рекламных объявлений может не иметь сопоставимой тяжести. Поэтому рекомендуется записывать отдельные поля «тяжесть при реализации», «качество сведений», «распространение», «срочность» и «доступность защиты». Любая итоговая категория применяется только для внутренней сортировки аналитической работы.

При пересмотре различаются изменения исходной системы, изменение экспозиции пользователей и изменение представления исследователя. Если новая проверка опровергла первоначальную гипотезу, карточка помечается соответствующим образом. Удалять её без следа неправильно: история таких ошибок необходима для оценки устойчивости метода и предотвращения повторных необоснованных обвинений.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Сценарий	Система плюс событие плюс последствие

Элемент проверки	Доказательственный критерий
Шкала	Тяжесть отдельно от уверенности
Остаток	После учёта подтверждённых мер
Эскалация	Необратимый вред и уязвимые лица

Ключевые источники для сопоставления: [8], [14], [17], [19].

Практический разбор 20.18

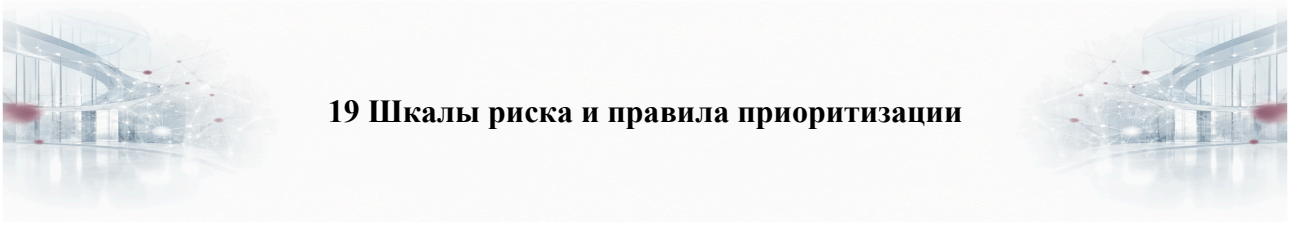
Рекомендация: Связывать риск с мерой и владельцем

Условный пример: В реестре записан риск ошибочного объединения людей без действия.

Разбор: Запись сама по себе не снижает вероятность вреда.

Практические действия: Владелец вводит проверку различающих идентификаторов и тестирует её на условных однофамильцах.

Критерий результата: У риска есть ответственная мера и подтверждение её проверки.



19 Шкалы риска и правила приоритизации

Прозрачные уровни

Внутреннюю шкалу рекомендуется ограничить четырьмя уровнями действий: обычное наблюдение, усиленная проверка, приоритетное исследование и срочная эскалация. Уровень определяется не только количественным счётом, но и характером возможного вреда: лишение свободы, неправомерное вмешательство в семейную жизнь, жизненно важная медицинская помощь и невозможность оспаривания способны требовать внимания даже при единственном кейсе. Показатели «низкий» и «высокий» не должны превращаться в оценку личности заявителя или доверия к организации.

Матрица принятия решений

Для каждой карточки фиксируется сочетание тяжести, убедительности источника, числа затрагиваемых лиц, срочности и возможности предотвратить ущерб. При срочности можно начать защитные действия до полного исследования, однако публичное заявление должно отдельно отражать степень подтверждения. Например, запрос о сохранении электронных логов может быть разумным при обоснованной угрозе их уничтожения, но это не даёт права утверждать, что виновность доказана.

Пересмотр приоритетов

Ежеквартально редакция оценивает, не стала ли система приоритетов несправедливо благоприятной для наиболее медийных стран и коммерчески известных разработчиков. Отдельно проверяется, какие группы и языки представлены недостаточно. Пересмотр критериев осуществляется до выбора новых дел и публикуется в методическом журнале; задним числом менять веса ради получения нужного результата нельзя. Для значимых изменений следует пересчитать исторические агрегаты.

Механизм неотложного приоритета и пересмотра

Для приоритизации важна система предохранителей от «красных флагов» в прессе. Громкая публикация может привлечь внимание, но высокий приоритет должен определяться

потенциальным вредом, возможностью его предотвращения и качеством информации. Предлагается двухэтапная оценка: первичный триаж угрозы человеку, затем независимая проверка доказательственного основания. Если срочное действие необходимо, оно может ограничиваться конфиденциальным запросом о сохранении логов без публичного заявления о виновности.

Долгосрочно следует анализировать долю проектов по регионам и группам заявителей. Если почти все исследования касаются англоязычных рынков, нельзя называть программу глобальным мониторингом без оговорки. Исправление смещения достигается не искусственным повышением балла отдельным странам, а планированием языковых ресурсов и новых каналов источников.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Низкий уровень	Периодическое обновление
Повышенный	Уточнение документов и контекста
Высокий	Независимая техническая проверка
Срочный	Протокол сохранения доказательств и эскалации

Ключевые источники для сопоставления: [8], [9], [14].

Практический разбор 20.19

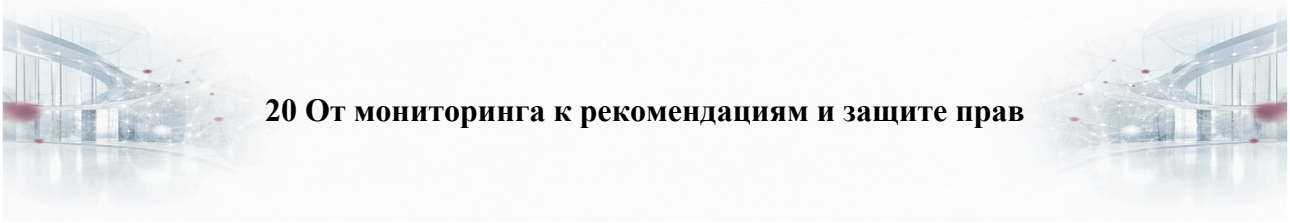
Рекомендация: Обосновывать приоритет по последствиям

Условный пример: Популярный медиасюжет вытесняет менее заметное продолжающееся ограничение прав.

Разбор: Частота упоминаний не равна тяжести и срочности.

Практические действия: Координатор оценивает подтверждённость, тяжесть, обратимость и продолжающийся характер.

Критерий результата: Очередность рассмотрения объяснена по единой шкале.



20 От мониторинга к рекомендациям и защите прав

Рекомендация не является предписанием

Observatory может предложить разработчику раскрыть документацию, оператору — пересмотреть процедуру, регулятору — проверить вопрос в рамках компетенции, а исследовательскому сообществу — разработать тесты. Однако нельзя оформлять это как обязательное распоряжение. Рекомендации делятся на правовые, технические, организационные и исследовательские. Для каждой обозначаются адресат, предмет, подтверждающее наблюдение, ожидаемый результат и условие последующей проверки.

Учет ответов

Наличие ответа организации не тождественно исполнению рекомендации. Поэтому отдельно фиксируются четыре статуса: принято к рассмотрению, заявлено о действии, действие подтверждено документом, эффект подтверждён независимыми данными. Если организация сообщает об исправлении, Observatory не снимает вывод без проверки, применимой к

характеру системы. В противном случае общественный мониторинг станет каталогом обещаний участников рынка.

Права затронутого лица

Для людей, пострадавших от автоматизации, полезнее всего указать доступные официальные средства защиты: запрос информации, обращение к контролёру данных, административную жалобу, судебное оспаривание, представительство адвоката. Доклад может содержать общую навигацию, но не обещает юридическую победу и не заменяет профессиональное консультирование в конкретной юрисдикции. Возможность пересмотра решения должна оцениваться по реальной доступности, срокам и затратам, а не по формальному наличию электронной формы.

Проверка исполнимости адресной рекомендации

Рекомендации целесообразно делить на действия, которые адресат может выполнить самостоятельно, и меры, требующие компетентного органа. Например, исправление неверной анкеты может находиться у оператора системы, а отмена юридически значимого ограничения — зависеть от банка, суда или регулятора. Observatoire должен указывать цепь компетенций и не обещать заявителю правовой эффект собственной публикации. В чувствительных делах общие рекомендации о порядке жалобы не заменяют персональной адвокатской работы.

Для оценки выполнения рекомендуется сохранять промежуточные статусы и документы: получение рекомендации, согласие рассмотреть, утверждённый план действий, фактическое внедрение и результат независимой проверки. Отсутствие подтверждения означает «эффект не установлен», а не «рекомендация провалилась» или «система не изменилась». Честная типология препятствует использованию мониторинга как списка формальных обещаний.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Публичная рекомендация	Не создавать видимость обязательного решения
Заявленное исправление	Отдельно от независимого подтверждения
Практическая защита	Проверяемый официальный путь обжалования
Дальнейший мониторинг	Дата повторной оценки

Ключевые источники для сопоставления: [3], [11], [13], [18].

Практический разбор 20.20

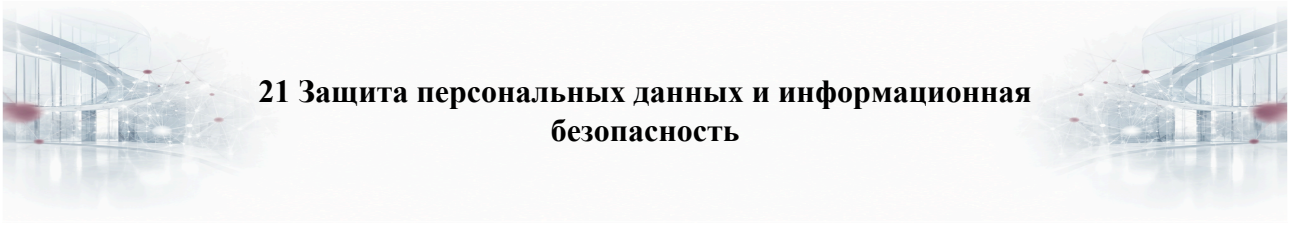
Рекомендация: Давать рекомендацию с исполнимым действием

Условный пример: Отчёт рекомендует повысить прозрачность решения об отказе.

Разбор: Адресат не понимает, какую запись или процедуру изменить.

Практические действия: Автор предлагает показывать исходные данные и основание отказа, называет владельца и способ проверки.

Критерий результата: Рекомендация содержит действие, адресата и критерий исполнения.



21 Защита персональных данных и информационная безопасность

Правовая оценка обработки

При обработке данных о конкретных лицах Observatory должен определить роль контролёра или обработчика, цели и основание обработки, категорию сведений, получателей и сроки хранения. В ЕС GDPR требует проверки принципов минимизации, точности, ограничения цели и безопасности. Особые категории данных, сведения о правонарушениях и данные несовершеннолетних могут требовать дополнительных условий и специальных гарантий. Сведения в открытых источниках не становятся автоматически свободными от правовой защиты; публичная доступность — лишь одно из обстоятельств анализа.

Для каждого набора персональных данных до приёма необходимо определить роль организации (контролёр или обработчик), конкретные цели, основание по статье 6 GDPR, сроки хранения и адресатов. Общественно полезная исследовательская цель сама по себе не отменяет требований GDPR. Если сведения затрагивают здоровье, политические убеждения, биометрическую идентификацию или другие специальные категории, требуется также применимое условие статьи 9. Исключение для научных исследований нельзя объявлять универсальным: его условия и национальные гарантии должны быть установлены для конкретной обработки. [13]

Особый юридический барьер представляет статья 10 GDPR. Сведения об уголовных осуждениях, правонарушениях и связанных мерах безопасности, которые могут встречаться в материалах международного розыска, нельзя свободно переносить в открытый реестр на основании общего интереса к мониторингу. Их обработка требует контроля официального органа или предусмотренного правом ЕС либо государства-члена разрешения и надлежащих гарантий. До подтверждения соответствующего режима Observatory следует ограничиться обезличенными структурными индикаторами и публичными решениями с проверкой допустимости дальнейшего воспроизведения; даже общедоступность документа не устраняет требований GDPR. [13]

Перед запуском канала индивидуальных обращений оцениваются статья 14 GDPR (информирование при получении данных не от субъекта), статья 28 (договоры с обработчиками), статья 32 (безопасность), статья 35 (DPIA при вероятном высоком риске), статьи 33–34 (реагирование на нарушения безопасности) и статьи 44–49 (передача данных за пределы ЕЭЗ). Применимость каждой обязанности определяется фактической обработкой; упоминание положения здесь не является выводом, что обязанность обязательно возникает в каждом случае. Псевдонимизация снижает риск, но не делает идентифицируемые данные автоматически анонимными. [13]

Вопрос допуска	Обязательная проверка перед обработкой
Обычные персональные данные	Цель, роль, основание статьи 6, минимизация и срок хранения
Специальные категории	Применимое исключение статьи 9 и национальные гарантии

Данные о правонарушениях	Отдельная проверка допустимости по статье 10 и применимому национальному праву
Высокий риск для прав	Оценка необходимости DPIA и меры снижения остаточного риска
Передача вне ЕЭЗ	Правовой механизм передачи, получатели и документированные гарантии

Разделение сред

Рекомендуется отдельно хранить контактные данные заявителей, исследовательский корпус и публичные агрегаты. Доступ выдаётся по ролям, проверяется через журналы, а экспорт материалов подлежит одобрению. Для передачи файлов используются защищённые каналы; прикрепление полного досье к обычному письму не должно быть стандартной практикой. При привлечении внешнего эксперта заранее определяются допустимые операции, запрет повторного использования и порядок удаления данных после заключения.

Оценка утечки

При нарушении безопасности важно не только выполнить применимые требования об уведомлении, но и не увеличить вред поспешной публикацией. План реагирования включает изоляцию доступа, сохранение доказательств, оценку типов раскрытых данных, возможное уведомление компетентного органа и пострадавших в предусмотренных правом случаях. Общественный отчёт о самой деятельности Observatory не должен скрывать значимых собственных ошибок организации. Ежегодно проверяется минимизация данных, а не только сложность технической защиты.

Правовые основания, специальные данные и безопасность

Структура законной обработки должна соответствовать специфике Observatory. Исследователь может законно анализировать опубликованный судебный акт, но при объединении разных баз существенно повышается риск идентификации лица и вторичного использования данных в ином контексте. Поэтому каждое новое связывание массивов требует отдельной оценки цели и необходимости. Нельзя хранить полный паспорт клиента в карточке алгоритмического риска только ради удобства последующего поиска.

Рекомендуется проводить тест на повторную идентификацию перед выпуском открытых данных: какие сведения доступны третьему лицу и можно ли по ним восстановить личность при помощи публичного поиска? При высоком риске часть атрибутов агрегируется или не публикуется. Внутренний журнал раскрытия данных помогает восстановить, какие версии и кому передавались, если впоследствии обнаружено нарушение безопасности.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Основание	Обосновывать до сбора личных сведений
Доступ	По ролям, с регистрацией выдачи
Хранение	Ограничивать целью и сроком
Утечка	Отдельная реакция и оценка правовых уведомлений

Ключевые источники для сопоставления: [13], [18].

Практический разбор 20.21

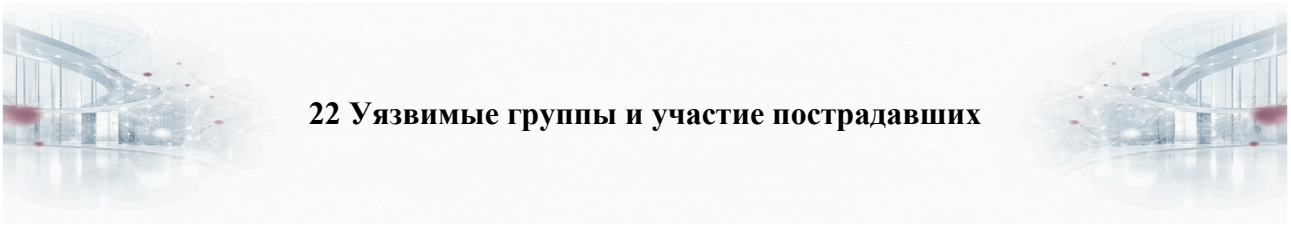
Рекомендация: Проверять косвенную идентификацию

Условный пример: Из карточки удалили имя, но сохранили редкую профессию и точную дату.

Разбор: Сочетание признаков позволяет узнать заявителя.

Практические действия: Ответственный оценивает необходимость деталей, обобщает их и ограничивает доступ к оригиналу.

Критерий результата: Открытая версия не раскрывает лишних идентифицирующих сочетаний.



22 Уязвимые группы и участие пострадавших

Невидимый вред

Последствия автоматизации не всегда видны в агрегированной статистике. Для мигрантов, лиц с ограниченной цифровой доступностью, несовершеннолетних или граждан в сложной финансовой ситуации неудачное решение может иметь непропорциональный эффект. Observatory должен оценивать не только вероятность ошибки, но и способность затронутого лица узнать о ней, получить объяснение и исправить результат. В отсутствие этих процедур даже сравнительно редкий алгоритмический сбой может создавать глубокое неравенство.

Осмысленное участие

Участие пострадавших не сводится к заполнению анкеты в конце проекта. При разработке методики необходимо учитывать понятность языка, безопасное обращение, недоверие к учреждениям и риск повторной травматизации. Экспертная группа может привлекать представителей затрагиваемых сообществ к оценке последствий, но не превращает их рассказы в доказательство технического механизма без дополнительной проверки. Вознаграждение за экспертное участие должно быть прозрачным и не зависеть от нужного вывода.

Правила публикации

Систематический мониторинг уязвимых групп требует строгой анонимизации или иной законной модели защиты данных. Удаление имени недостаточно, если совокупность даты, местности и редких обстоятельств позволяет узнать человека. При конфликте между информативностью текста и безопасностью заявителя приоритет получает предотвращение необоснованного раскрытия. Дело может быть представлено в виде агрегированного сценария с ясно обозначенным изменением деталей и сохранением доказательств в закрытом архиве.

Участие затронутых лиц без вторичного вреда

Качественное участие затронутых людей требует самостоятельного планирования. Перевод формы жалобы на другой язык не устраняет ограничения доступа для человека без документов, средств связи или свободного времени. Исследовательская команда должна изучать реальные барьеры обращения и не путать малое число жалоб с отсутствием вреда. Для уязвимых групп могут потребоваться посредники, соблюдающие конфиденциальность и не получающие право менять свидетельство заявителя.

При разработке метрик следует учитывать тяжесть единичного вреда. Если одна автоматизированная ошибка лишает человека жизненно важной услуги, простая средняя «точность модели 99%» не описывает достаточность защиты. Для такого случая значимы своевременность исправления и наличие резервного человеческого механизма. Публикуя

историю, Observatory должен обеспечить защиту не только имени, но и редкой комбинации социальных обстоятельств.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Канал обращения	Доступные язык и формат
Оценка ущерба	Учитывать обратимость и процессуальное положение
Ко-дизайн	Участие без давления на вывод
Публикация	Тест на повторную идентификацию

Ключевые источники для сопоставления: [3], [4], [13], [18].

Практический разбор 20.22

Рекомендация: Учитывать доступность процедуры участия

Условный пример: Форма жалобы доступна только на одном языке и требует сложного файла.

Разбор: Наблюдение систематически пропускает часть затронутых лиц.

Практические действия: Команда вводит понятные альтернативные каналы и учитывает смену охвата в показателях.

Критерий результата: Доступность улучшена, сравнение периодов скорректировано.

23 Международный обмен данными и трансграничные кейсы

Согласование общей структуры карточки не означает разрешение обмениваться исходными файлами. Международным коллегам сначала направляются методические определения и обезличенные показатели; передача исходных индивидуальных материалов требует отдельного анализа роли получателя, правового основания и требований к трансграничной передаче. Доступ международной организации, научного эксперта или консультанта нельзя считать автоматически освобождённым от правил защиты данных. [13]

Передача и локализация

Наблюдение за системой, работающей в нескольких странах, может включать персональные данные, банковскую тайну и материалы уголовных производств. Для каждой трансграничной передачи необходимо установить участников, цель, законность первоначального получения и основания дальнейшей обработки. В европейском контексте могут применяться положения главы V GDPR; в иных юрисдикциях действуют свои требования. Общий интерес к глобальной сравнительной аналитике не отменяет ограничений юрисдикции происхождения материалов.

Правило минимального раскрытия

Международная рабочая группа не должна по умолчанию получать одинаковый доступ ко всему архиву. Доступ целесообразно организовывать по распределённой модели: локальный хранитель проверяет первичные материалы, а международным участникам передаёт минимальный набор подтверждённых атрибутов, пригодных для сопоставления. Если цель можно достигнуть статистическими показателями без пересылки досье, выбирается такой

вариант. Это особенно важно для сообщений о политически чувствительных преследованиях и международном розыске.

Сопоставимость процедур

Зарубежный документ может подтверждать факт задержания, но не законность основания розыска; судебный отказ в выдаче — решение по конкретной процедуре, но не универсальное доказательство отсутствия международных полицейских записей. Observatory обязан сохранять различие между состоянием процедуры и оценкой правового режима. Для сравнения кейсов вводятся поля «документ», «орган», «дата», «процессуальный эффект», «юрисдикция» и «доказательное ограничение».

Трансграничный обмен: минимизация и условия передачи

Для трансграничной работы целесообразно использовать модель «локального хранителя»: первичные личные данные остаются у правомерного держателя, а международная команда получает минимальный обезличенный набор признаков. Это снижает риск несовместимых национальных правил и избыточной передачи персональных данных. Если вывод невозможно проверить без прямого доступа к документу, назначается компетентный независимый проверяющий на месте с ограниченными полномочиями. Публичное заключение описывает уровень доступа рецензента, не раскрывая тайну.

Особую сложность представляет несовпадение процессуальных статусов. Судебное прекращение одного производства не всегда означает прекращение другого в иной стране. Автоматизированная запись в базе и судебное решение также не равнозначны. При построении международного профиля дела необходимо хранить отдельные линии компетенций, не превращая хронологию в упрощённый «глобальный статус».

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Получение материала	Законная цель и происхождение
Доступ коллег	Минимальный обоснованный набор
Публичная карточка	Только проверяемые нейтральные атрибуты
Трансграничный акт	Не обобщать эффект на все юрисдикции

Ключевые источники для сопоставления: [11], [13], [18].

Практический разбор 20.23

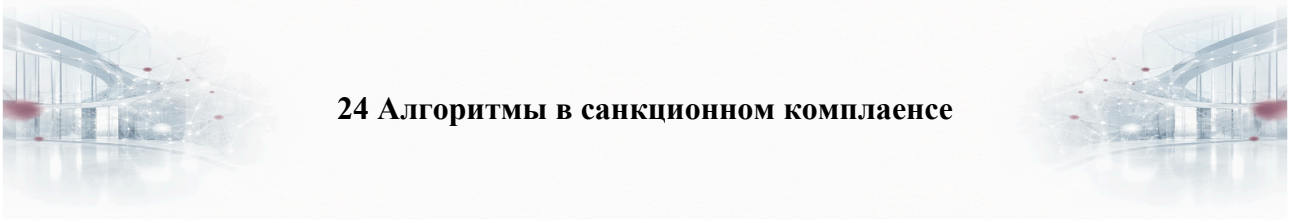
Рекомендация: Минимизировать международный обмен

Условный пример: Партнёр запрашивает полный архив для проверки одного эпизода.

Разбор: Объём запроса превышает исследовательскую задачу.

Практические действия: Координатор выделяет релевантные материалы, проверяет основание и защищённый маршрут передачи.

Критерий результата: Передача документирована по составу, цели и получателю.



24 Алгоритмы в санкционном комплаенсе

Особенности сектора

Санкционный скрининг обычно сочетает правила совпадения имён, транслитерацию, признаки собственников и оценку связей. Ошибка может привести к блокировке операции, отказу в услуге и длительному сбору документов. Однако само совпадение с именем не доказывает ошибку или незаконность решения: значение имеют действующие санкционные режимы, сведения идентификации и внутренние обязанности организации. Observatory наблюдает не только техническую точность, но и наличие понятного процесса пересмотра, сроки ответа и долю случаев, где отказ мотивирован соответствующими правовыми основаниями.

Метрики без ложной объективности

Показатель false positive rate имеет смысл только при установленной процедуре определения истинного статуса и известном наборе отрицательных примеров. Если независимой разметки нет, допустима лишь метрика «доля обращений, признанных ошибочными по итогам пересмотра» — она измеряет процесс пересмотра, а не истинную точность скрининга. Систематическое исключение сложных имён из тестовой выборки приводит к смещённой оценке. Протокол должен отдельно учитывать транслитерацию, порядок имён, географические варианты и обновления списков. [14]

Правозащитная составляющая

Особое внимание уделяется доступности механизма исправления идентификации и сохранению письменной истории взаимодействия. Исследователь не вправе обещать снятие санкций или доступ к счёту на основании собственной оценки; эти решения находятся у компетентных субъектов. Публичный отчёт может обосновать рекомендации по процедурам уведомления и апелляции, не раскрывая персональные банковские досье и не выдавая проектную схему контрольных мер за обязательное требование для каждого банка.

Верификация ложных совпадений без презумпции ошибки

Проверка санкционного скрининга должна учитывать, что цена ложноположительного сигнала и цена пропущенного действительного совпадения различаются, но ни одна из них не отменяет юридическую обязанность организации. Для исследовательского отчёта особенно важно отделять техническую метрику модели от правовой оценки решения банка. Например, повторяющееся совпадение имён может требовать настройки процедуры человеческой верификации, однако не является доказательством необоснованного отказа без анализа документов.

Для условной тестовой выборки формируется набор имён с контролируемыми вариантами транслитерации, псевдонимами и различиями идентификаторов. Тест проводится без реальных сведений клиентов, если возможна синтетическая модель. Отчёт ограничивает область вывода испытанным набором и открыто показывает случаи, в которых не удалось установить ground truth. Предложения по прозрачности формулируются так, чтобы не раскрывать чувствительные пороги санкционного контроля.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Совпадение имени	Не равно установленной идентичности
Ошибка системы	Нужна независимая ground truth
Метрика пересмотра	Отдельно от технической FPR
Защита клиента	Документирование, жалоба, человеческая проверка

Ключевые источники для сопоставления: [6], [11], [13], [14].

Практический разбор 20.24

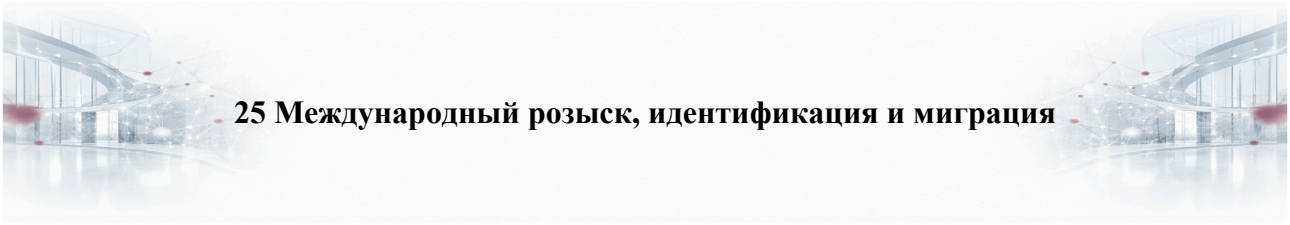
Рекомендация: Проверять санкционное совпадение отдельно от риска

Условный пример: Карточка описывает отказ однофамильцу как установленное санкционное лицо.

Разбор: Машинный флаг и идентичность не тождественны.

Практические действия: Эксперт сверяет доступные идентификаторы и сохраняет спорное совпадение как вопрос проверки.

Критерий результата: Публикация не приписывает лицу неподтверждённый санкционный статус.



25 Международный розыск, идентификация и миграция

Условия повышенного риска

Автоматизация установления совпадений личности, анализа документов и приоритизации запросов может иметь особенно серьёзные последствия, если результат используют в пограничном контроле или международном розыске. Observatory не имеет доступа к закрытым базам INTERPOL по умолчанию и не должен заявлять, что способен подтвердить индивидуальный статус розыска по публичным материалам. Исследование концентрируется на проверяемых институциональных гарантиях: качества сопоставления, сохранения причин решения, человеческой проверки, механизмов исправления и соблюдения процессуальных прав.

Цепочка решения

Алгоритмическая подсказка, полицейская запись, национальный ордер, задержание и решение об экстрадиции — разные этапы. Ошибкой будет приписать весь итог работе модели либо смешать международный обмен данными с судебной компетенцией государства. Для правомерного анализа нужно документировать, какой орган принял каждое решение, какой именно цифровой инструмент использовался и что могло изменить результат. В отсутствие технических логов выводы о механизме системы остаются предварительными, даже если последствия для человека подтверждены.

Публичное раскрытие

Дела об убежище, преследовании и международной полицейской информации часто включают угрозу безопасности заявителя. Публикация имени, миграционного маршрута и биометрических характеристик может усугубить риск. Поэтому отраслевые отчёты должны агрегировать процессуальные наблюдения или использовать условные композитные сценарии. Рекомендации адресуются принципам точности и доступа к эффективному средству защиты, а не публичному определению «кто находится в розыске».

Недопустимость вывода о роли ИИ без доказательств

Исследовательская карта международного розыска должна строго различать техническое совпадение, внесение информации в полицейскую систему, национальную меру принуждения и судебное решение об экстрадиции. Эти процессы могут относиться к разным правопорядкам и не образуют единого автоматического юридического статуса. Наличие цифровой обработки не доказывает, что именно AI-система создала основание задержания. Для причинного вывода требуются документы об использованном инструменте и о последующем решении компетентного лица.

Для публичного кейса достаточно рассмотреть структуру гарантий без идентификации человека: существование процедуры исправления, независимость рассмотрения жалобы, сохранение оснований решения и возможность адвокатского сопровождения. Если доступны только сообщения адвоката и судебное постановление, техническая часть маркируется как неизвестная. Это не делает пережитый вред несущественным, но защищает анализ от приписывания недоказанной технологической причины.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Тип решения	Отделять сигнал, меру и судебный акт
Источник	Только законно доступный документ
Причинная связь	Не предполагать применение ИИ без доказательства
Персональные риски	По умолчанию обезличенный анализ

Ключевые источники для сопоставления: [3], [10], [11], [13], [18].

Практический разбор 20.25

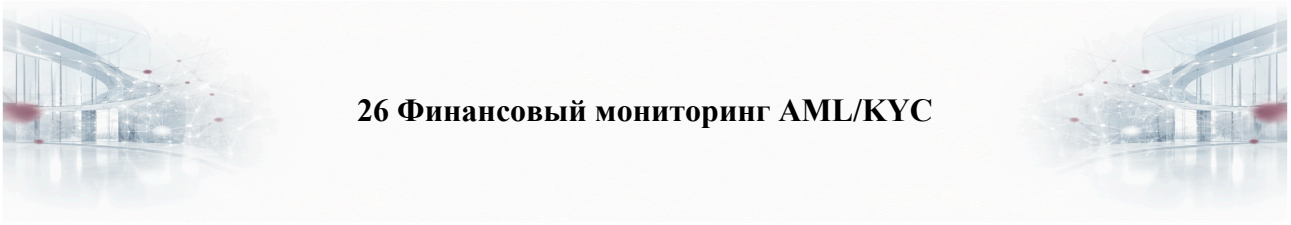
Рекомендация: Разделять официальную запись и автоматическую сводку

Условный пример: Алгоритм объединил публикации о разных людях в одну розыскную историю.

Разбор: Сводка не подтверждает официальный статус и может смешивать лиц.

Практические действия: Юрист связывает каждое утверждение с первичным материалом и исключает необоснованные объединения.

Критерий результата: В карточке раздельно указаны источник, лицо и доказанный статус.



26 Финансовый мониторинг AML/KYC

Процесс принятия решения

При AML/KYC автоматизированные инструменты выявляют аномалии, формируют предупреждения и помогают аналитикам изучать финансовые операции. Наличие предупреждения само по себе не доказывает преступления и не означает окончательного отказа клиенту. Observatory должен исследовать прозрачность правил человеческой эскалации, доступность корректировки неверных данных и качество внутренних механизмов разграничения риска и санкции. Конфиденциальность противодействия отмыванию денег при этом может ограничивать информацию, доступную клиенту и внешнему исследователю.

Состав метрик

Для анализа полезны несколько несводимых показателей: доля предупреждений, эскалированных аналитиком; доля закрытых без дополнительных действий; длительность ручного пересмотра; изменение нагрузки; пропорция протестированных кейсов из разных групп клиентов. Эти показатели не должны интерпретироваться как «эффективность выявления преступлений», если отсутствуют подтверждённые основания. Нельзя публиковать число направленных сообщений в органы финансовой разведки как оценку преступности без системной оговорки о природе таких сообщений.

Рекомендации

Приоритетом становятся процессы качества идентификации, правила периодического пересмотра моделей, разделение обязанностей и сохранение достаточной информации для законной проверки. Публичный мониторинг может анализировать стандарты защиты прав и агрегированные результаты, не раскрывая критерии, позволяющие обходить финансовый контроль. Технические тесты должны избегать распространения реальных персональных данных и банковских инструкций, которые могли бы использоваться во вред.

Границы вывода об AML-скоринге и отказе в услуге

В AML/KYC полезна модель трёх решений: алгоритм формирует предупреждение, аналитик рассматривает его, уполномоченная организация принимает юридически значимую меру. Каждый шаг может иметь свои источники ошибки и не должен объединяться в недифференцированную «автоматическую блокировку». Мониторинг проверяет качество данных, критерии направления на дополнительную проверку, документирование человеческого участия и эффективность процедуры исправления неверной идентификации. Конфиденциальные механизмы выявления подозрительных операций требуют осторожности при публичном раскрытии.

При исследовании агрегированных показателей важно установить рабочие единицы: доля предупреждений, доля закрытых обращений, срок рассмотрения и число реально оспоренных решений. Без данных о подтверждённых противоправных действиях нельзя назвать число предупреждений «пойманными преступлениями». Для независимого сравнения банков следует отмечать различия продуктовой линейки, клиентской базы и объёма операций, иначе статистика будет наказанием за более широкий охват контроля.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Алгоритмический сигнал	Не установленный финансовый проступок
Эскалация	Доля по известному объёму предупреждений
Процедурная защита	Срок пересмотра и документируемый ответ
Публикация	Не раскрывать чувствительные контрольные правила

Ключевые источники для сопоставления: [6], [11], [13], [14].

Практический разбор 20.26

Рекомендация: Не приравнивать AML-сигнал к преступлению

Условный пример: Заявитель получил отказ после автоматической оценки риска.

Разбор: Факт отказа не устанавливает нарушение закона клиентом.

Практические действия: Аналитик описывает подтверждённое действие и запрашивает допустимые основания пересмотра.

Критерий результата: Текст сохраняет различие между риском, решением и юридической квалификацией.

27 Государственные услуги и автоматизация решений

Публичная функция

При распределении пособий, постановке в очередь или оценке нуждаемости применение ИИ способно влиять на гарантированные права. Исследование не должно ограничиваться точностью классификатора. Нужны сведения о нормативном основании решения, роли должностного лица, понятности уведомления, фактической возможности обжалования и последствиях задержки исправления. Контекстная оценка прав человека помогает выявить процессы, где даже небольшая ошибка становится серьёзным препятствием для уязвимого человека.

Открытость без иллюзий

Реестр государственных AI-проектов полезен, но сам факт записи не подтверждает, что проект действует в заявленном виде. Observatory сопоставляет сведения о закупке, эксплуатации, независимых проверках и регуляторных решениях. Для международного сравнения учитывается разница между пилотной программой и масштабным внедрением. Отдельно фиксируются проекты, завершённые после обнаружения проблем, чтобы показатели не создавали ошибочную картину бесконечного роста автоматизации.

Альтернатива автоматизации

Обоснованная оценка может прийти к выводу, что задача вообще не требует ИИ. В соответствии с подходом риск-менеджмента целесообразно сравнивать алгоритмическое решение с существующей процедурой по качеству, стоимости, доступности объяснения и правовым последствиям. Общественный интерес требует не максимизации числа внедрений, а выбора процедуры, которая обеспечивает права и достижение цели. Доказательная база для такого сравнения должна учитывать ресурсные ограничения учреждений.

Контроль мотивировки решения и доступного обжалования

Карта публичной услуги начинается с вопроса о цели автоматизации. Если организация оптимизирует скорость принятия решений, но не измеряет долю неправомерных отказов и доступность исправления, эффективность оценивается неполно. Для Observatory практическая единица — путь человека от подачи заявления до решения и возможного пересмотра. В этом пути фиксируются точки, когда алгоритмический прогноз превращается в фактическое последствие, кто может изменить решение и какой документ остаётся у заявителя.

Общество выигрывает от доступной критики алгоритмических решений, но сама по себе публикация кода не обеспечивает понятной процедуры защиты. Поэтому мониторинг должен оценивать связку «объяснение — возможность представления возражения — компетентный пересмотр — реальное исправление». При сравнении учреждений учитывается социальная задача и доступность нецифровой альтернативы для лиц, которые не могут использовать онлайн-сервис.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Правовая основа	Компетенция и критерии административного решения
Объяснение	Понятное человеку основание
Обжалование	Практическая доступность и сроки
Альтернатива	Сравнение с неавтоматизированной процедурой

Ключевые источники для сопоставления: [3], [9], [11], [14].

Практический разбор 20.27

Рекомендация: Проверять доступность пересмотра публичного решения

Условный пример: В уведомлении об отказе в услуге нет понятного способа исправить данные.

Разбор: Техническое описание модели не решает проблему заявителя.

Практические действия: Команда устанавливает оператора, исходные данные и предусмотренный маршрут оспаривания.

Критерий результата: Рекомендация адресована владельцу процедуры и проверяется учебным обращением.



28 Закупки ИИ и контракты публичного сектора

Место договоров

В международном мониторинге закупка является одним из ключевых доказательств внедрения. Контракт может раскрывать обязанности по журналированию, обновлению модели, безопасности и использованию данных. Однако закупочная запись не означает технического развёртывания. Observatory различает объявление конкурса, заключение договора, ввод в эксплуатацию, изменения, приёмку и прекращение. Если поставщик сохраняет контроль над моделью, особенно важны условия доступа государственного заказчика к документам и последствиям инцидентов.

Права на проверку

Предлагается рекомендовать закупщикам условия о предоставлении проверяемой документации независимому оценщику, уведомлении об изменениях, хранении логов, раздельной ответственности и возможности приостановить применение. Эти положения остаются проектными рекомендациями, а не универсальными действующими требованиями. В коммерчески чувствительных случаях допускается закрытая проверка с публичным описанием области тестирования, существенных недостатков и ограничений доступа.

Сопоставление расходов и выгод

Сравнение контрактов должно учитывать цену приобретения, технической поддержки, миграции данных, обучения персонала и расходы на исправление ошибочных решений. Публичная цена подписки без дополнительных обязательств не описывает стоимость владения. Если контракт засекречен, запрещено приписывать ему условия на основании типового предложения компании. Вывод обозначается как ограниченный открытыми сведениями, а рекомендации направляются на прозрачность условий контроля.

Проверяемость закупочных и договорных обязательств

Закупочный мониторинг может отслеживать сроки проекта как последовательность отдельных статусов, а не как один показатель «AI внедрён». Применяются отметки конкурса, заключения договора, пилота, приёмки, рабочей эксплуатации и прекращения. Аналитик проверяет, содержит ли контракт возможность независимо получить логи, документацию и сведения об обновлениях при возникновении претензий. Даже если таких условий нет, нельзя задним числом заявлять о нарушении, пока не установлена применимая правовая норма.

Для общественной аналитики полезен минимальный набор контрактных рекомендаций: устойчивый идентификатор системы, обязанность уведомлять о существенном изменении, меры реагирования на инциденты, распределение ответственности за качество данных и механизм независимого рассмотрения жалоб. Эти требования следует обозначать как модельные условия закупки. При сравнении стоимости владения необходимы расходы на обслуживание, обучение, миграцию данных и процедуру исправления ошибок, а не только подписка на программу.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Закупка	Подтверждённая стадия, не просто анонс
Обновление	Договорные обязанности уведомления
Проверка	Ограниченный доступ независимого оценщика
Стоимость	Полный жизненный цикл

Ключевые источники для сопоставления: [3], [11], [14], [16].

Практический разбор 20.28

Рекомендация: Включать проверяемые условия в закупку

Условный пример: Публичный заказчик требует этичный ИИ без критериев приёмки.

Разбор: Общее пожелание не позволяет проверить исполнение.

Практические действия: Эксперт предлагает версии, журнал, независимый доступ и тест спорного решения как предмет приёмки.

Критерий результата: Контрактный контроль связан с конкретными свидетельствами.



29 Frontier AI и модели общего назначения

Объект мониторинга

Модель общего назначения может быть базовым компонентом тысяч приложений, поэтому её оценка требует отдельной конструкции. Observatory описывает модель, версию, доступность весов, набор известных тестов, границы доступа и области применения. Результаты разработчика не считаются независимым подтверждением безопасности. Международный AI Safety Report 2026 подчёркивает разрыв между результатами лабораторных тестов и реальным поведением систем, что особенно важно для агентных функций и новых возможностей.

Системные риски

Риски включают масштабируемые ошибки, злоупотребления, безопасность инфраструктуры и влияние на общественные системы. Они не сводятся к индивидуальному событию. Для мониторинга полезны регулярные изменения тестовых результатов, независимость оценщика, документирование мер защиты и наличие процедуры реагирования. Но единичный закрытый тест не позволяет доказать универсальную безопасность модели. В отчёте отдельно показываются непротестированные области, потенциальная утечка тестовых данных и влияние выбора запросов.

Правовой статус

EU AI Act устанавливает отдельные обязанности для поставщиков GPAI-моделей и более строгие меры для моделей с системным риском. Их нельзя автоматически переносить на всех пользователей любого чат-бота. Observatory анализирует обязанности по актуальной норме, роли субъекта и дате действия, а собственные рекомендации маркирует как добровольные. Выводы о техническом потенциале модели не должны представляться решением о нарушении закона без полномочий и полной фактической базы.

Оценка рисков GPAI без подмены статуса регулятора

Мониторинг frontier AI требует явного различия между оценкой возможностей и выводом о вреде. Результат теста способен подтвердить, что система решила определённый набор задач, но не автоматически показать, как она ведёт себя в неизвестной среде и как злоумышленник мог бы её использовать. Доклад должен фиксировать конфигурацию, версию, доступные инструменты, ограничения доступа и тестовый протокол. Особенно опасна ссылка на «экспертный бенчмарк» без возможности понять его репрезентативность.

Для моделей общего назначения предлагается реестр обновлений результатов независимых тестов, отдельных сообщений о реальных инцидентах и заявлений разработчиков о защитных мерах. Эти слои не сводятся в общую оценку безопасности. Если конкретные технические детали закрыты, Observatory публикует рамку проверки, а не имитирует полный аудит. Вопрос юридических обязанностей GPAI анализируется отдельно от научного вопроса об измеренной способности модели.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Модель	Версия, провайдер и доступ
Тест	Независимость и протокол
Риск	Отдельно возможности и последствия
Закон	Категория GPAI и применимая обязанность

Ключевые источники для сопоставления: [11], [12], [15], [19].

Практический разбор 20.29

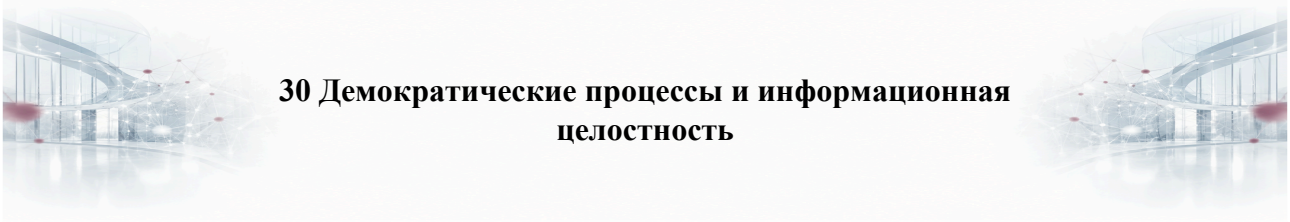
Рекомендация: Не переносить общий рейтинг на развёрнутый продукт

Условный пример: Высокий рейтинг модели используют для обоснования автономных действий помощника.

Разбор: Рейтинг не охватывает права инструментов и контекст оператора.

Практические действия: Оценщик запрашивает паспорт конфигурации и профильные испытания действий.

Критерий результата: Вывод относится к фактически проверенному режиму применения.



30 Демократические процессы и информационная целостность

Риски публичной сферы

Генеративные модели могут использоваться для создания убедительных ложных материалов, имитации голоса и масштабирования манипулятивных сообщений. При изучении таких явлений необходимо отличать автоматизированную генерацию от сознательного обмана, а документированную кампанию — от предположения об организаторе. Особенно высок риск ложных обвинений в отношении политических и общественных деятелей. Observatory должен придерживаться нейтральных процедур, опубликованных до исследования, и исключать редакционное влияние партийных интересов.

Показатели

Подсчёт синтетических материалов зависит от качества идентификации. Детектор AI-generated текста не является непреложным доказательством происхождения, особенно при переводе и редактировании. Поэтому публикуются уровни уверенности, методы ручной проверки и доля спорных объектов. Показатели охвата должны учитывать платформы и языки; нельзя механически выводить общественное влияние из числа найденных публикаций. Причинная связь с поведением аудитории требует отдельного исследования.

Права и баланс

Противодействие манипуляциям не должно превращаться в механизм подавления законного выражения мнений. Рекомендации формулируются с учётом свободы выражения, приватности и разумных ограничений доказательств. Если выявлена ошибка в собственной оценке происхождения контента, исправление публикуется с такой же заметностью, как первоначальное утверждение. Для чувствительных выборных периодов предусмотрен усиленный этический контроль, а не ускоренное снижение стандарта проверки.

Отграничение доказанного вмешательства от гипотез

Наблюдение за информационной целостностью должно учитывать фундаментальную проблему атрибуции. Даже качественный детектор синтетического изображения или текста не доказывает организацию кампании и личность автора. Для вывода о координации нужны дополнительные данные, которые часто закрыты. Observatory обязан сообщать о технических ограничениях обнаружения, не публикуя необоснованные персональные обвинения. Само сообщение о предполагаемой манипуляции может усилить распространение недостоверного материала.

Процедура публикации включает оценку общественного интереса и пропорциональности. Нужен ли полный повтор изображения или достаточно обезличенного описания приёмов? Не повредит ли повторная публикация участнику? Для серьёзных выводов полезна внешняя проверка специалистом по цифровой криминалистике и правам человека. Возражение затронутой стороны рассматривается по тем же стандартам, независимо от её политического статуса.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Контент	Отдельно происхождение и истинность
Детектор	Оценка ошибки и спорные случаи
Влияние	Не выводить из просмотров причинность
Нейтральность	Отвод политически заинтересованных рецензентов

Ключевые источники для сопоставления: [3], [6], [9], [19].

Практический разбор 20.30

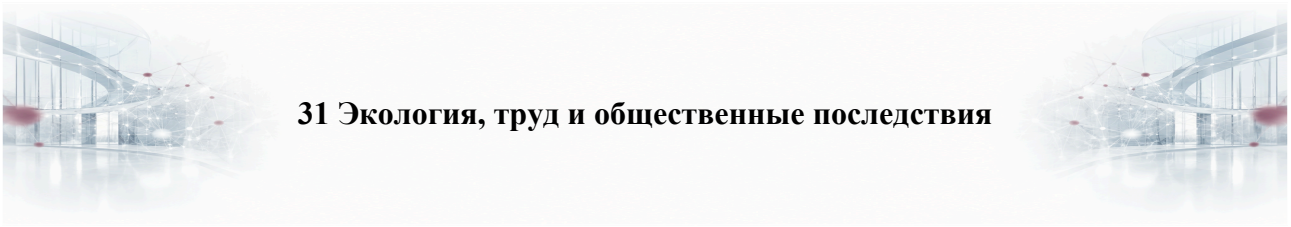
Рекомендация: Отделять происхождение контента от его влияния

Условный пример: Материал предположительно создан ИИ и широко распространяется.

Разбор: Происхождение, достоверность и воздействие требуют разных доказательств.

Практические действия: Исследователь проверяет каждый вопрос отдельно и сохраняет неопределённость атрибуции.

Критерий результата: Публикация не подменяет доказательство влияния предположением об авторстве.



31 Экология, труд и общественные последствия

Системный масштаб

Независимое наблюдение должно включать не только наиболее драматичные инциденты, но и распределённые эффекты: нагрузку вычислительной инфраструктуры, условия труда при разметке данных, доступность технологий и неравенство ресурсов. Такие показатели редко привязаны к одному событию. Для оценки воздействия нужны сопоставимые методики, географический контекст, виды энергии и границы системы. Публикация единственного числа «углеродного следа ИИ» без объяснения состава расчёта приводит к недобросовестным сравнениям.

Трудовые процессы

Системы автоматической оценки работников могут влиять на условия труда, продвижение и дисциплинарные решения. Observatory обращает внимание на информирование, человеческий пересмотр, репрезентативность оценочных критериев и жалобы работников. Важно отличать тестирование алгоритма от доказательства дискриминационного решения в конкретной организации. Случаи изучаются с соблюдением защиты персональных данных и с возможностью безопасного обращения представителей работников.

Метод сопоставления

При оценке экологических и социальных эффектов рекомендуется публиковать границы учёта и чувствительность выводов к допущениям. Оценка по доступным отчётам производителя маркируется как self-reported, а независимая проверка выделяется отдельно. Публичный dashboard не должен смешивать техническую энергоэффективность с распределением социальных выгод. Международное сравнение может честно завершаться выводом о невозможности надёжного ранжирования в силу отсутствия открытых данных.

Оценка внешних эффектов и границ расчётов

В экологической оценке границы системы важнее эффектного итогового числа. Должно быть ясно, измеряется ли обучение модели, работа сервиса, инфраструктура дата-центра или жизненный цикл оборудования. Сравнение между организациями возможно только при одинаковом охвате и единицах. Данные разработчика маркируются как самодекларированные, а оценка внешнего исследователя — с указанием способа проверки. Отсутствие достоверных данных у одной страны не доказывает меньший экологический эффект.

При трудовом мониторинге важно отделять роль ИИ в автоматизированном ранжировании и самостоятельные решения работодателя. Групповое неравенство результатов требует статистической проверки и юридического контекста. Обобщение конкретной жалобы не должно раскрывать личность работника, если речь идёт о малом коллективе. Модель Observatory предусматривает отдельную оценку социальной доступности технологии, чтобы выгоды и риски рассматривались симметрично.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Энергия	Границы системы и период
Труд	Процедуры обжалования решений
Заявление компании	Отдельно от независимой проверки
Неизмеримость	Прямо отражать отсутствующие данные

Ключевые источники для сопоставления: [3], [6], [17], [19].

Практический разбор 20.31

Рекомендация: Сопоставлять общественные издержки по определённой базе

Условный пример: Два сервиса сравнивают по энергозатратам без одинаковой единицы нагрузки.

Разбор: Числа могут отражать разный объём и тип работы.

Практические действия: Аналитик задаёт сравнимую задачу, границы измерения и источники; при нехватке данных отказывается от рейтинга.

Критерий результата: Сравнение воспроизводимо либо ограничения явно препятствуют ранжированию.

32 Взаимодействие с заявителями и гражданским обществом

Структурное участие

Независимость не означает изоляции экспертов от людей, на которых воздействуют технологии. Observatory может создавать консультационные группы из исследователей, представителей общественных объединений, профессиональных сообществ и лиц с опытом алгоритмических решений. Участники должны знать, какие вопросы решаются консультационно, а какие находятся в компетенции научной редакции. Нельзя превращать участие гражданского общества в декоративное одобрение уже принятого вывода.

Предотвращение асимметрии

Крупные технологические организации имеют больше ресурсов на подготовку ответов и экспертиз, чем отдельные заявители. Чтобы не воспроизводить этот перекос, предусматриваются понятные формы подачи, перевод ключевых материалов и доступный обзор методики. Одновременно соблюдается право проверяемой организации представить документированное возражение. Процедура не обязана считать все мнения одинаково подтверждёнными, но должна объяснять, почему доказательственная сила материалов различается.

Представительство

Членство в экспертном совете не должно считаться полномочием говорить от имени всей социальной группы. При подборе участников оцениваются разные страны, дисциплины и типы опыта. Для долгосрочных исследований полезно описывать не только присутствие «заинтересованных сторон», но и случаи, когда их замечания привели к изменению метода. Такой показатель менее эффектен, но показывает реальное качество обратной связи.

Процедура консультаций и защита интересов участников

Участие общественности должно быть устроено так, чтобы мнения действительно могли изменить метод. На раннем этапе участникам предлагают обсуждать определения вреда, категории уязвимости, каналы сообщения и язык уведомлений, а не утверждать готовые выводы. Для этого нужен журнал замечаний с ответами: принято, принято частично, отклонено по обоснованной причине. В публичном отчёте достаточно обобщить, как участие изменило протокол без раскрытия личности заявителей.

Следует предотвращать захват консультаций наиболее ресурсными группами. Участникам предоставляется понятный доступ к материалам, равные основные сроки и возможность представлять информацию без сложных технических форматов. Экспертный совет в свою очередь не обязан признавать любое сообщение доказанным; независимость подразумевает справедливую процедуру оценки фактов, а не голосование за научный результат.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Включение	Разные дисциплины и группы
Роль	Консультация отдельно от утверждения
Ответ	Задокументировать учёт возражений

Элемент проверки	Доказательственный критерий
Эффект	Примеры изменений методики после участия

Ключевые источники для сопоставления: [1], [3], [4], [18].

Практический разбор 20.32

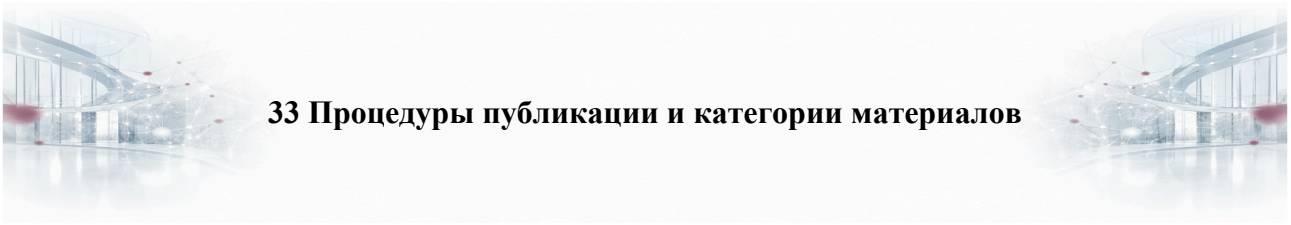
Рекомендация: Согласовывать ожидания заявителя

Условный пример: Человек ожидает, что публикация автоматически отменит отказ банка.

Разбор: Исследовательская работа не гарантирует индивидуального решения.

Практические действия: Координатор объясняет этапы проверки, возможные результаты и отдельный маршрут защиты прав.

Критерий результата: Заявитель понимает статус участия и способы дальнейшего обращения.



33 Процедуры публикации и категории материалов

Типы выпусков

Рекомендуется четыре класса публикаций: оперативная аналитическая записка, тематический мониторинговый отчёт, методический стандарт и ежегодный институциональный обзор. Оперативная записка может быть краткой, но обязана указывать статус проверки и обновлений; тематический доклад раскрывает выборку и допущения; стандарт описывает применимую методику; ежегодный обзор фиксирует также собственные ошибки Observatory. Название жанра предотвращает ситуацию, когда предварительный сигнал представляют как окончательный научный результат.

Редакционный стандарт

Каждый документ содержит дату среза, авторов или редакционную ответственность, источники, метод включения событий, правовые ограничения, определение ключевых понятий, право ответа и историю изменений. При серьёзной неопределённости соответствующая оговорка появляется в резюме, а не скрывается в примечании в конце. Для публикаций, предназначенных регуляторам или судам, дополнительно фиксируется, какие утверждения основаны на первичных документах и какие остаются экспертной интерпретацией.

Открытые данные

Раскрытие машиночитаемых данных желательно, но только после правовой и этической оценки. В реестре могут публиковаться обезличенные коды, даты и подтверждённые категории без распространения персональных досье. Метаданные включают версию классификатора, историю обновлений и условия повторного использования. Если открытая публикация недопустима, необходимо объяснить ограничение и предоставить статистические агрегаты либо процедуру контролируемого научного доступа, когда она возможна.

Уровни доказательств и условия публикации

Статусы публикации предлагается обозначать явно: «предварительное наблюдение», «проверенный исследовательский отчёт», «методическая позиция» и «исправленная версия».

Это облегчает работу прессы и юридических адресатов, поскольку одни материалы предназначены для раннего выявления риска, а другие — для обоснованного сравнительного вывода. В оперативном материале особенно важно не скрывать неопределённость ради краткости. Заголовок не должен утверждать больше, чем подтверждает основная часть.

Публикация открытых данных допускается после независимого теста на раскрытие личной информации и проверки условий вторичного использования. Полный технический лог, банковская переписка или материал индивидуального дела не превращаются в открытую базу потому, что они интересны исследователям. При невозможности раскрыть первичные документы допустимо опубликовать описания процедуры проверки и ограниченный набор агрегированных данных, достаточный для независимого понимания вывода.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Сигнал	Предварительная записка с предупреждением
Обзор	Метод выборки и ограничения
Стандарт	Версия методики и отзыв рецензентов
Данные	Лицензия и защита персональной информации

Ключевые источники для сопоставления: [7], [8], [13], [14].

Практический разбор 20.33

Рекомендация: Публиковать материал по степени подтверждённости

Условный пример: Срочный сигнал хотят выпустить в формате окончательного расследования.

Разбор: Формат создаёт избыточную уверенность.

Практические действия: Редактор выбирает предварительное сообщение с вопросами проверки и условиями обновления.

Критерий результата: Заголовок и содержание соответствуют доказательственному статусу.

34 Управление исправлениями, отзывами и оспариванием

Право на исправление

Организация должна предусматривать процедуру оспаривания публикации: заявитель указывает конкретный тезис, представляет основание возражения и способ связи. Редакция подтверждает получение, определяет серьёзность риска, назначает независимого проверяющего и принимает мотивированное решение. Исправление может заключаться в обновлении факта, изменении степени уверенности, уточнении правового статуса или отзыве фрагмента. Оспаривание не должно превращаться в право проверяемого субъекта остановить нежелательное исследование.

Версионирование

Тихая замена PDF без истории изменений подрывает прослеживаемость, особенно если исследование цитируют в юридических документах. Каждый отчёт получает номер версии, журнал исправлений и устойчивую ссылку. Если ошибка серьёзна, старая версия отмечается как заменённая, а в новой указывается причина. Исключение допускается для случаев, когда сохранение ошибочной версии само по себе раскрывает опасные персональные данные; тогда история описывается без сохранения незаконного содержимого.

Независимый пересмотр

При конфликте между автором публикации и заявителем запрос о повторной проверке передаётся специалисту, не участвовавшему в первоначальном исследовании. Вынесенное решение фиксирует оценку доказательств и остаточную неопределённость. Количество удовлетворённых жалоб следует публиковать вместе с количеством всех обращений и разъяснением процедурных критериев. Высокая доля исправлений может означать как плохое качество первой проверки, так и прозрачную способность исправлять ошибки; оба объяснения требуют изучения.

Исправление ошибки с сохранением следа версий

Система исправлений требует разграничения технической опечатки, фактической ошибки, несогласия с интерпретацией и нового последующего события. Их нельзя считать одинаковыми «корректировками». Например, изменение закона после публикации не означает, что первоначальная юридическая характеристика была ошибочной на дату её составления; требуется обновление актуальности. Напротив, неверно названный человек требует исправления ошибочного факта и оценки последствий для его репутации.

При материальной ошибке редакция уведомляет известных непосредственных пользователей заключения, когда это возможно и соразмерно. Версия сохраняет номер, дату и перечень существенных изменений. В процедуре оспаривания отсутствует автоматическое право на удаление критического, но проверенного суждения; заявителю гарантируются рассмотрение аргументов и мотивированный ответ, а редакции — обязанность исправить установленный недостаток.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Обращение	Указать оспариваемый тезис и основание
Триаж	Приоритет при риске непосредственного вреда
Версия	Дата, содержание и причина исправления
Апелляция	Независимый от автора пересмотр

Ключевые источники для сопоставления: [13], [14], [18].

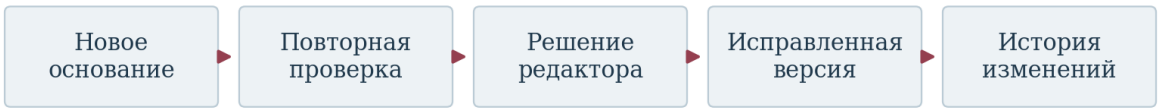


Рисунок 20.4. Исправление публичной карточки. Процедура раздела 34. Исправление сохраняет историю версий и доступное объяснение изменения существенного вывода.

Практический разбор 20.34

Рекомендация: Сохранять историю исправлений

Условный пример: После получения нового документа редактор молча меняет вывод.

Разбор: Читатели прежней версии продолжают использовать ошибочный текст.

Практические действия: Ответственный выпускает датированное исправление и связывает его с прежней версией.

Критерий результата: Пользователь видит, что изменено, почему и какая версия актуальна.

35 Контроль качества исследований и peer review

До публикации

Перед публикацией проверяется соответствие исследовательского вопроса, источников, метода и вывода. Методологический рецензент должен восстановить основной расчёт и оценить альтернативное объяснение; предметный эксперт — проверить смысл классификации; юрист — применимость правовых утверждений. Редактор фиксирует существенные замечания и основания решения по ним. Работа обозначается как прошедшая внешнюю научную рецензию только при фактическом выполнении соответствующей процедуры.

Разделение проверок

Независимое кодирование выполняется до совместного обсуждения. Исходные оценки сохраняются отдельными записями; согласованная категория — последующее решение с мотивировкой. Межэкспертная согласованность характеризует воспроизводимость классификации, а не истинность первоначального сообщения. Поэтому рядом с её оценкой проверяются качество источников, полнота материалов и пригодность самого определения события. [M1; M3]

Послепубликационный контроль

После публикации выборка карточек проходит повторную проверку от источника до вывода. Отбор для оценки общей доли ошибок и целевой отбор наиболее чувствительных карточек выполняют раздельно. В первом случае описываются вероятности включения и неопределённость; во втором результаты используются для выявления существенных дефектов без заявления о частоте во всём корпусе. Исправления возвращаются в классификатор и процедуру сбора.

Независимость рецензии и межэкспертная согласованность

Рецензент получает исследовательский вопрос, правила включения, словарь категорий, доступные первичные материалы и расчёт. Его задача — установить, следует ли вывод из процедуры и данных, а не подтвердить заранее выбранную позицию редакции. Конфликт интересов и пределы доступа указываются до работы. Для повторной проверки сохраняются замечания и ответы, включая отклонённые предложения с мотивировкой.

Независимая повторная разметка должна включать обычные, пограничные и недостаточно документированные случаи. После разбора ошибок инструкция уточняется, а её пригодность проверяется на новых материалах. Использование того же набора после совместного согласования оценивает обучение на примерах, но не переносимость правила. Основное внимание уделяется расхождениям, меняющим смысл публичного вывода или риск для затронутого лица.

В учебном наборе из 40 карточек два независимых оценщика совпали в 32 случаях и каждый отнёс 20 карточек к положительной категории. Наблюдаемое согласие равно 0,80, согласие по произведениям предельных долей — 0,50, поэтому к Коэна = (0,80 − 0,50) / (1 − 0,50) = 0,60. Это условный расчёт, а не результат ARGA. Коэффициент интерпретируется вместе с таблицей расхождений, распределением классов и характером спорных случаев; универсальный порог пригодности методики из него не выводится. [M3]

Разбор согласованности независимого кодирования

Условный расчёт к из раздела 35 основан на следующей совместной таблице. Категории обозначают наличие или отсутствие заранее определённого признака события; принадлежность к положительной категории не означает юридическую виновность.

Оценщик A	B положительно	B отрицательно
Положительно	16	4
Отрицательно	4	16

Согласие наблюдается в 16 + 16 = 32 из 40 карточек. Каждый оценщик использовал положительную категорию в половине случаев, поэтому базовое ожидаемое согласие по

предельным долям равно $0{,}5^2 + 0{,}5^2 = 0{,}50$. Полученное к = 0,60 сопровождается разбором восьми расхождений. Если они связаны с отсутствием первичного документа, повышение согласия требует лучшего материала; если с определением категории — уточнения инструкции. [M3]

После согласования сохраняются исходные метки и отдельная окончательная категория. Пересчёт согласия только по совместно исправленным ответам не оценивает первоначальную воспроизводимость. Новая инструкция проверяется на свежем наборе. Для редкой категории или другого распределения классов коэффициент может вести себя иначе, поэтому вместе с ним всегда публикуются исходные количества и содержательное значение разногласий.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Редактор	Язык, структура, аккуратность
Научный рецензент	Метод и устойчивость вывода
Юрист	Применимость нормы и риски раскрытия
Аудитор данных	Происхождение, расчёт, воспроизводимость

Ключевые источники для сопоставления: [9], [14], [16].



Рисунок 20.3. Согласованность независимого кодирования. Условные данные раздела 35: совпадения 16 + 16, расхождения 4 + 4. Наблюдаемое согласие 80%, ожидаемое 50%. Расчёт не является результатом ARGA.

Практический разбор 20.35

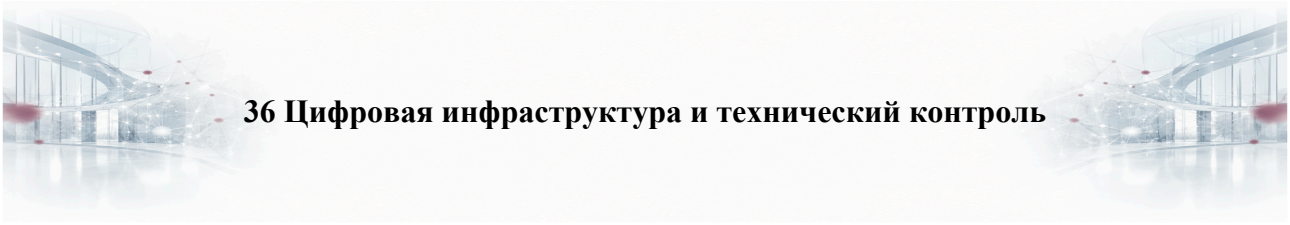
Рекомендация: Проводить независимую рецензию ключевых выводов

Условный пример: Автор одновременно выбрал источники и утвердил причинный вывод.

Разбор: Необнаруженное смещение может пройти весь процесс.

Практические действия: Куратор передаёт рецензенту источники, правила отбора и альтернативы, сохраняет ответы на замечания.

Критерий результата: Существенные расхождения разрешены либо отражены как ограничения.



36 Цифровая инфраструктура и технический контроль

Архитектура данных

Практический Observatory требует минимальной инфраструктуры: отдельного реестра систем, реестра событий, хранилища доказательств, журналов доступа, системы задач и портала публикаций. Связи между реестрами выражаются устойчивыми идентификаторами без дублирования персональных данных. Публичная часть не должна иметь прямого доступа к закрытому доказательственному архиву. Все изменения карточек сопровождаются журналированием автора, времени и основания, а удаление осуществляется с учетом процедур сохранения юридически значимых материалов.

Автоматизация работы

Инструменты ИИ могут помогать обнаруживать дубликаты, переводить документы, извлекать метаданные и классифицировать сообщения, но окончательная оценка вреда и статуса должна проходить человеческую проверку. Использование генеративной модели в анализе документируется: указываются задачи, версии инструмента, процедуры контроля ошибок и ограничения передачи данных. Нельзя незаметно отправлять закрытые материалы заявителей во внешние AI-сервисы без соответствующей правовой и технической оценки.

Устойчивость

Минимальные меры включают резервное копирование, многокомпонентную аутентификацию, управление ключами, ограничение экспорта и регулярную проверку восстановления. Система должна выдерживать отказы поставщика, так как потеря архива может разрушить доказательственную цепочку. Для публичных данных желательна возможность независимого скачивания версий, а для конфиденциальных — поимённый допуск и запись выдачи. План безопасности пересматривается после инцидента или значимого обновления инфраструктуры.

Защита архива и воспроизводимость вычислений

Техническая инфраструктура должна обеспечивать проверяемую историю, а не только удобный интерфейс. Публикуемый идентификатор карточки связывается с архивом доказательств через закрытую таблицу соответствия, что позволяет исправлять публичную запись без распространения персональных сведений. Всякий экспорт требует проверки схемы: нет ли в метаданных спрятанных авторов, адресов, координат или исходных названий файлов. Для высокочувствительных данных применяются ограниченный доступ и отдельные правила резервного копирования.

При автоматизации отбора событий неизбежны ошибки классификации. Поэтому AI-инструмент должен выдавать кандидаты и объяснимую для человека запись, а окончательное решение подтверждает аналитик. Использование одного и того же генеративного

инструмента для создания сообщения и его независимой проверки не гарантирует независимости. В периодических тестах учитываются новые версии модели и возможные изменения качества перевода или извлечения.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Реестр	Версии и неизменяемый журнал действий
Архив	Локальное отделение закрытых материалов
Публикация	Проверка персональных атрибутов перед экспортом
Автоматизация	Только при документированной человеческой верификации

Ключевые источники для сопоставления: [13], [14], [15].

Практический разбор 20.36

Рекомендация: Проверять восстановление и разграничение доступа

Условный пример: Архив резервируется, но восстановление никогда не испытывалось.

Разбор: Наличие копии не подтверждает возможность вернуть рабочие данные.

Практические действия: Администратор восстанавливает учебный набор в отдельном контуре и проверяет роли.

Критерий результата: Тест восстановления и отказа неразрешённого доступа завершён протоколом.

37 Метрики деятельности самой обсерватории

Качество вместо PR

Деятельность Observatory оценивается по свойствам исследовательского процесса: проверяемости происхождения данных, воспроизводимости классификации, своевременности исправлений и сохранению существенных ограничений в публикации. Число выпущенных материалов характеризует объём работы. Оно не измеряет достоверность корпуса или уменьшение общественного вреда. Для каждого показателя фиксируются источник, знаменатель и возможный способ его искусственного улучшения.

Результативность

Практический эффект оценивается через подтверждённые изменения процедур, использование рекомендаций в публичных документах, качество реакции адресатов и доступность механизмов защиты заявителей. Цитирование отчёта не доказывает, что поведение организаций изменилось из-за него. Для причинных выводов нужно учитывать другие реформы и временные факторы. В ежегодном отчёте можно обозначать «наблюдаемый вклад», но не автоматически заявлять собственную заслугу в государственном или судебном решении.

Контрметрики

Для предотвращения искажения целей полезно публиковать показатели отказов от недостаточно подтверждённых материалов, числа запросов о корректировке и доли случаев, где независимая проверка не подтвердила предварительную гипотезу. Это не провал

исследовательской организации, а проявление методической дисциплины. Руководство должно вознаграждать качество отказа от неубедительного вывода так же, как качество доказанного результата.

Разделение показателей выпуска и общественного эффекта

Для институционального отчёта полезно заранее установить противоположные пары показателей. Число опубликованных карточек сопровождается долей карточек, отклонённых из-за слабых источников; число рекомендаций — долей с подтверждённой обратной связью; скорость публикации — числом материальных исправлений. Такой набор сдерживает стимул к завышению объёмов. Руководство не должно вознаграждать отдельного исследователя за количество скандалов или постов без проверки качества.

При сравнении работы по годам учитываются изменения штата и доступности источников. Рост числа обнаруженных событий может отражать развитие самого Observatory. Для международного KPI необходимо публиковать активные языки, страны, типы источников и долю подтверждённых записей. Если бюджет на независимое рецензирование сократился, это обстоятельство должно отражаться в годовом отчёте как ограничение качества.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Качество источников	Доля карточек с первичной опорой
Независимость	Доля процедур с раскрытым отводом
Ошибка	Срок исправления обоснованной претензии
Влияние	Подтверждённая реакция адресата, без присвоения заслуг

Ключевые источники для сопоставления: [8], [14], [20].

Практический разбор 20.37

Рекомендация: Измерять качество собственной работы

Условный пример: Команда отчитывается количеством карточек, не считая исправления.

Разбор: Рост выпуска может скрывать ухудшение проверки.

Практические действия: Аналитик учитывает существенные ошибки, сроки исправления и повторную рецензию по определённой выборке.

Критерий результата: Отчёт показывает качество и ограничения оценки наряду с объёмом.

38 Сотрудничество с международными организациями

Модели участия

Для независимой экспертной структуры возможны законные формы участия в публичных консультациях, тематических обсуждениях, научных мероприятиях и исследовательских сетях. Они не означают автоматической аккредитации в ООН или статуса органа ОЭСР. Любое заявление о сотрудничестве должно указывать фактическую форму: направленный комментарий, подтверждённое участие, совместная публикация или институциональное соглашение. Смешение форм наносит вред доверию к научной независимости.

Соответствие мандатам

При работе с UNESCO полезны страновые методики RAM и инструмент EIA, а с OECD — единые определения инцидентов и принципы AI governance. Для научной группы ООН актуальны обобщённые данные о рисках и исследовательские методы, но каналы подачи определяются официальной процедурой соответствующего органа. Observatory не должен направлять персональные материалы в международный форум только потому, что вопрос имеет трансграничный характер. Предпочтительнее аккуратная агрегированная аналитика с методическим приложением.

Совместные исследования

Потенциальные партнёры могут согласовывать словари данных, независимое рецензирование, методики наблюдения и двуязычные аннотации. Договор устанавливает владение результатами, право публикации, защиту данных и процедуру разрешения научных разногласий. Совместный логотип недопустим без реального согласия другой организации. Публичный сайт должен содержать проверяемые доказательства заявленных связей: официальные письма, страницы мероприятий или ссылки на совместные материалы.

Разграничение взаимодействия и институционального признания

Перед заявлением о взаимодействии с международной организацией предлагается внутренний тест документальной подтверждаемости. Что имеется: письмо-приглашение, регистрация участника, подтверждение получения исследования, официальный совместный проект или только факт цитирования методики? Каждое обстоятельство требует собственной формулировки. Направленный комментарий не следует превращать в заявление об официальном одобрении содержания. Сведения о статусе проверяются по публичным страницам соответствующих организаций, когда такие страницы существуют.

Совместный методический проект должен заранее разграничивать владение исходными данными, публикационную автономию, порядок возможных исправлений и право прекратить сотрудничество. Для международных форумов, где принимаются добровольные экспертные материалы, полезнее предоставить короткую проверяемую записку с приложенной методикой, чем большой отчёт с неопределённым качеством доказательств.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
ООН	Фактический канал участия, не подразумеваемый статус
UNESCO	Сопоставление RAM/EIA
OECD	Инциденты и интероперабельность метрик
Договор	Сохранение независимых редакционных решений

Ключевые источники для сопоставления: [1], [2], [3], [4], [5], [7], [8], [21], [22].

Практический разбор 20.38

Рекомендация: Подтверждать институциональные заявления документами

Условный пример: Подтверждение получения комментария называют официальным партнёрством.

Разбор: Административный ответ не подтверждает такой статус.

Практические действия: Редактор сверяет формулировку с письмом и выбирает точное описание участия.

Критерий результата: Каждое заявление о сотрудничестве имеет соответствующее основание.



39 Пилотная программа Observatoire ARGA

Пилот вместо декларации

Пилот начинается с одной ведущей предметной области и ограниченного набора источников, чтобы ошибки метода можно было отделить от различий между секторами. Для ARGA подходящим исследовательским предметом могут стать документированные случаи ошибочной идентификации при санкционном скрининге, если доступны законные и проверяемые материалы. Границы языка, территории, периода и каналов фиксируются до сбора. Два других профильных направления остаются кандидатами на расширение после оценки первой серии.

Этапы

Программа включает разработку словаря и правил отбора, пробную разметку учебных карточек, сбор ограниченного корпуса, независимое кодирование нового поднабора и выпуск методического отчёта. Каждая стадия имеет проверяемый результат. До приёма чувствительных материалов проверяются законность обработки, права доступа и восстановление архива. Продолжительность стадий определяется ресурсным решением, а не объявляется фактически утверждённым календарём.

Критерии продолжения

Переход к расширению требует воспроизводимого пути от источника к карточке, объяснённых разногласий классификации, работающего исправления и достаточных ресурсов защиты данных. Если эти условия не выполнены, определяется конкретная доработка или сужение наблюдаемой области. Пилот может завершиться обоснованным отказом от сравнения стран или расчёта частоты при сохранении полезности документированных индивидуальных случаев.

Пилотный запуск: этапы и критерии остановки

До сбора регистрируются вопросы пилота: воспроизводится ли отбор; различают ли оценщики основные категории; можно ли проследить источник; как влияет появление нового канала; сколько времени занимает проверка и исправление? Численные критерии принимаются с учётом последствий ошибки и ресурсов, до знакомства с итоговыми результатами. Их изменение требует мотивированной записи. Неблагоприятный результат пилота сохраняется как основание улучшения методики.

На решение о продолжении влияют как положительные, так и отрицательные сигналы. Если не удалось защитить персональные сведения или независимую проверку, программа останавливает чувствительный канал, даже если интерес общества высок. В отчёт пилота включаются требования к расширению: новый язык, новый сектор, дополнительные рецензенты, бюджет на безопасность и подтверждённая потребность в источниках.

До начала пилота следует разделить три уровня доступа: открытые агрегированные факты без идентифицируемых данных; ограниченные рабочие карточки с проверенными основаниями; особо защищённые индивидуальные материалы, приём которых не начинается до правового согласования и тестирования информационной безопасности. Такое деление

предотвращает ситуацию, когда технически готовый сайт ошибочно воспринимается как разрешение хранить чувствительные досье. Переход между уровнями оформляется отдельным протоколом и фиксируется в реестре решений.

Итоговый отчёт пилота содержит поток включения и исключения событий, совместную таблицу независимого кодирования, причины существенных расхождений, анализ чувствительности к источникам и фактическую нагрузку проверки. Отдельно проверяются безопасность хранения и процедура исправления. На этой основе принимается решение о следующем языке, секторе или канале. Отсутствие критической находки в небольшой учебной серии не трактуется как доказательство безопасности всей инфраструктуры.

Этап	Критерий продолжения / остановки
До открытия приёма	Роли, GDPR-проверка, меры безопасности и независимое утверждение процедуры
Тестовый набор	Прослеживаемое происхождение материалов и зафиксированные ограничения
Независимое кодирование	Оценена согласованность, спорные категории описаны и метод пересмотрен
Публикационный пилот	Сработало право на ответ, есть журнал изменений и понятная политика исправлений
Масштабирование	Нет неустранённых критических конфликтов, нарушений безопасности и пробелов доступа

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Этап 1	Устав и источники денег
Этап 2	Непубличная тестовая выборка
Этап 3	Пилотный методический отчёт
Этап 4	Внешняя оценка и решение о масштабе

Ключевые источники для сопоставления: [3], [4], [7], [8], [14].

Практический разбор 20.39

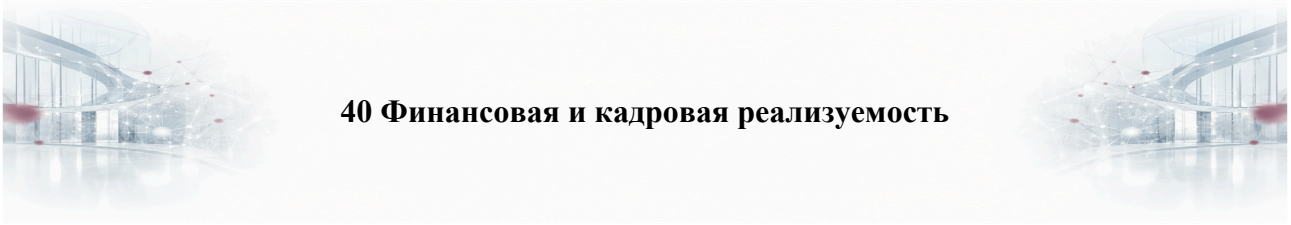
Рекомендация: Пилотировать ограниченный корпус до масштабирования

Условный пример: Команда планирует международный рейтинг до проверки кодирования.

Разбор: Несогласованные категории сделают сравнение произвольным.

Практические действия: Куратор проводит пилот на ограниченном корпусе, независимую разметку и анализ чувствительности.

Критерий результата: Решение о расширении основано на качестве процедуры и выявленных пробелах.



40 Финансовая и кадровая реализуемость

Функции важнее должностей

Бюджет следует формировать по функциям: исследование, правовая проверка, работа с данными, информационная безопасность, перевод и независимая рецензия. На первом этапе

функции допустимо совмещать, но не без конфликтных ограничений. Например, одно лицо может проводить сбор и первичное кодирование, однако не должно единолично утверждать собственный спорный вывод. Расходы на внешнюю правовую экспертизу и защиту обращающихся предусматриваются заранее, иначе организация будет вынуждена отказываться от наиболее общественно значимых кейсов.

Бюджетные сценарии

Поскольку реальные договоры и фонд оплаты труда для проектируемого центра в данном исследовании не представлены, нельзя выдавать условные суммы за подтверждённый бюджет ARGA. Рекомендуется использовать сценарии «минимальный», «базовый» и «расширенный» с диапазонами расходов после внутренней калькуляции. Каждый сценарий отдельно учитывает редакционную независимость, хранилище данных, языковую поддержку и резерв на реагирование. Экономия за счёт отсутствия защиты данных или рецензирования недопустима как устойчивый способ финансирования.

Управление дефицитом

В условиях нехватки ресурсов Observatory должен сокращать число наблюдаемых объектов, а не стандарт проверки. Публикации со слабой доказательной базой могут привлечь краткосрочное внимание, но увеличивают риск для затронутых лиц и доверия к организации. Отказ от чувствительного кейса по причине недоступности независимой проверки документируется без раскрытия заявителя. Сотрудники не должны зависеть от количества критических упоминаний или судебных последствий как показателя личной результативности.

Проверка ресурсной модели на операционных данных

Ресурсная модель полезна только при честном раскрытии неопределённости. Вместо произвольной суммы «годового бюджета международного Observatory» рекомендуется рассчитать стоимость конкретных операций: часов аналитика на одну карточку, времени юриста, перевода документа, независимого reviewer и защищённого хранения. Эти параметры сначала измеряются на пилоте и лишь затем масштабируются. Если затраты неизвестны, указывается «требуется калькуляция», а не вымышленная цифра.

При дефиците финансирования организация должна сохранять минимальный стандарт доказательств, сокращая число тем. Условный план расширения предусматривает отдельный резерв на корректировки, срочную защиту заявителей и технические инциденты. Это особенно важно для общественной организации, чья репутация зависит не столько от числа опубликованных документов, сколько от способности доказательно отвечать за каждый из них.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Исследование	План доступной аналитической мощности
Юридический контроль	Отдельный внешний бюджет
Информационная безопасность	Обязательная функция даже на пилоте
Резерв	Экспертиза и срочная защита данных

Ключевые источники для сопоставления: [14], [16], [18].

Практический разбор 20.40

Рекомендация: Соотносить обещания с кадровым ресурсом

Условный пример: Два сотрудника обещают ежедневную экспертизу всех поступающих дел.

Разбор: Нагрузка может исключить рецензию и своевременные исправления.

Практические действия: Руководитель оценивает время на карточку, вводит приоритеты и резерв на сложные случаи.

Критерий результата: План соответствует доступному времени с сохранением контроля качества.



41 Выводы и предлагаемая международная модель

Институциональный вывод

Научная состоятельность независимого AI Observatory определяется воспроизводимостью наблюдения и обоснованностью перехода к выводу. Контент-анализ обеспечивает единые правила классификации, выборочная методология раскрывает ограничения отбора, теория измерений задаёт смысл показателей. Институциональная независимость защищает возможность опубликовать неудобное наблюдение и исправить собственную ошибку. Совместное выполнение этих условий создаёт проверяемую исследовательскую функцию.

Приоритетные решения

Первоначально предлагается утвердить исследовательский протокол, словарь категорий, правила работы с источниками и конфликтами, режим защиты данных и процедуру исправлений. Затем провести пилот с независимым кодированием и анализом чувствительности. Международное сопоставление допускается лишь для показателей с обоснованной сопоставимостью определений и наблюдения. Формальное единство формы карточки не заменяет этой проверки.

Конечное назначение для ARGA

Практический путь ARGA — подтвердить работоспособность ограниченного наблюдения, затем расширять область по результатам проверки. Публикации должны сохранять различие между зарегистрированным сообщением, подтверждённым событием, исследованной причиной и оценкой распространённости. Материалы могут поддерживать консультации и защиту прав в пределах их доказательной силы; исследовательский статус не придаёт им полномочий суда или регулятора.

Проверяемые условия институционального внедрения

Перед масштабированием требуется проверить устойчивость отбора и кодирования на новом материале, влияние языкового расширения, ресурсную доступность рецензии и исполнение исправлений. При различиях между группами сначала рассматриваются различия наблюдаемости. Только после исключения или обоснованного учёта таких объяснений формулируется более широкий вывод. Неразрешённый пробел ограничивает конкретный показатель, а не скрывается в общем примечании.

Следующий этап исследования должен показать, какие элементы предложенной процедуры воспроизводятся разными специалистами и где требуется изменение метода. До выполнения пилота доклад является проектом научно обоснованной инфраструктуры. Его практическая ценность заключается в проверяемых критериях запуска, анализа и пересмотра,

позволяющих отделять подтверждённое наблюдение от более широкого утверждения, для которого данных пока недостаточно.

Контрольные вопросы и минимальные доказательства

Элемент проверки	Доказательственный критерий
Мандат	Исследование и публичный контроль
Процедура	Методы, данные, независимость
Вывод	Уровень уверенности и ограничения
Развитие	Пилот — проверка — масштабирование

Ключевые источники для сопоставления: [1], [2], [4], [6], [7], [9], [14].

Практический разбор 20.41

Рекомендация: Принимать решение о запуске по подтверждённой готовности

Условный пример: Реестр технически работает, но нет процедуры исправлений и ответственного за жалобы.

Разбор: Запуск увеличит последствия ошибок до готовности защитных процессов.

Практические действия: Куратор проверяет протокол, роли, защиту, рецензию и учебное исправление; незакрытые критические вопросы устраняет до запуска.

Критерий результата: Решение содержит подтверждения готовности, ограничения и условия пересмотра.

Практические приложения: рабочие формы и условные сценарии

Приложения являются проектными инструментами. Их поля и процедуры подлежат настройке с учётом применимого законодательства, характера доступа к данным и ресурсов организации. Приведённые сценарии не описывают реальных клиентов или установленные нарушения. Наличие формы не подтверждает её утверждения в качестве официального внутреннего регламента.

Приложение 1 Устав независимого AI Observatory

Назначение

Проект устава закрепляет общественно-аналитический характер организации. Основные задачи: собирать проверяемые сведения об использовании ИИ, разрабатывать методы измерения рисков, публиковать отчёты и содействовать содержательному международному обсуждению. Observatory не является государственным регулятором, органом сертификации или судебным учреждением и не может требовать раскрытия данных вне предусмотренных законом процедур. Устав должен иметь приоритет над коммерческими договорами, способными повлиять на редакционные решения.

Порядок применения и ограничения

Устав принимается компетентным органом организации после юридической проверки применимого права. Рекомендуемая формулировка: «Методологические выводы формируются на основании документированных исследований, а донор или заказчик не вправе определять результат оценки». Отдельно предусмотрены полномочия совета по назначению внешней рецензии, порядок пересмотра методик и независимая процедура работы с жалобами.

Поле/контроль	Правило заполнения или рекомендуемое значение
Предмет	Публичный научно-аналитический мониторинг ИИ
Пределы полномочий	Нет государственного принуждения или сертификации
Органы	Совет, секретариат, комитет этики, редакция
Независимость	Запрет на покупку или цензуру результатов
Публикация	Обязательная методика и история исправлений
Пересмотр	Ежегодная оценка устава и практики

Проверка на практике

Практический тест устава: специалист, получивший предложение провести исследование за счёт проверяемого разработчика, обязан проверить, может ли донор исключить критические выводы. При положительном ответе проект не объявляется независимым. Документ отдельно предусматривает публичный годовой отчёт об исправлениях и источниках финансирования. Пересмотр устава проводится по документированной процедуре; совет не может задним числом изменить правило ради конкретного кейса. При внешнем упоминании Observatory указывается действительное юридическое наименование организации и характер осуществляемой деятельности.

Приложение 2 Хартия независимости и редакционных гарантий

Назначение

Хартия применяется ко всем исследованиям, включая коммерчески оплачиваемые. Она запрещает финансирующей стороне право предварительного согласования выводов, право одностороннего удаления критических замечаний и требования скрыть сам факт существенного конфликта. Доступ к черновикам может предоставляться только для фактической проверки и защиты конфиденциальности с документированием полученных комментариев.

Обязательное раскрытие в проектной хартии: связь Observatoire с ARGA и иными организациями под общим управлением; категории платных услуг и доноров; порядок отвода при исследовании действующего или бывшего клиента; независимый внешний канал сообщения о давлении; запрет обусловливать оплату желательным выводом; процедуру исправления даже при неблагоприятном для самой ARGA результате. Пока эти гарантии не утверждены и фактически не проверены, термин «независимый» употребляется как цель институционального устройства, а не как доказанная сертифицированная характеристика.

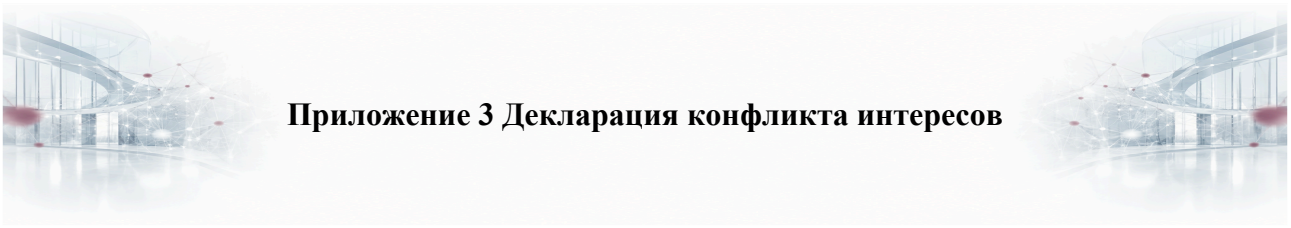
Порядок применения и ограничения

В пилотной практике каждый контракт должен проходить проверку на соответствие хартии. Нарушение ведёт к приостановке публикационного процесса, привлечению независимого юриста и возможному отказу от проекта. Публичная запись раскрывает факт применения хартии, но не публикует персональные данные сотрудников или законно защищённые договорные сведения без необходимости.

Поле/контроль	Правило заполнения или рекомендуемое значение
Заказчик	Не утверждает научный вывод
Эксперт	Раскрывает релевантные интересы
Рецензент	Не подчиняется проверяемой стороне
Право публикации	Ограничивается только законными основаниями
Возражения	Принимаются и учитываются, но не дают право вето
Санкция	Отвод или прекращение несовместимого задания

Проверка на практике

В договорном приложении к хартии рекомендуется прямо закрепить: заказчик получает право представить фактические замечания, но не утверждает научное заключение. При конфликте между договором конфиденциальности и общественно значимой информацией юрист определяет законно допустимый вариант публикации: полный, обезличенный, агрегированный или отказ. Неполучение права на раскрытие исходных данных не позволяет объявить проверку полноценным открытым аудитом. Документирование такого ограничения должно быть условием каждого итогового отчёта.



Приложение 3 Декларация конфликта интересов

Назначение

Форма заполняется перед каждым назначением и обновляется при изменении обстоятельств. В ней нельзя требовать чрезмерные сведения о семье и частной жизни; существенны только отношения, способные повлиять на объективность. Декларация содержит период последних профессиональных и финансовых связей, участие в разработке проверяемой системы и возможное владение интересом конкурента.

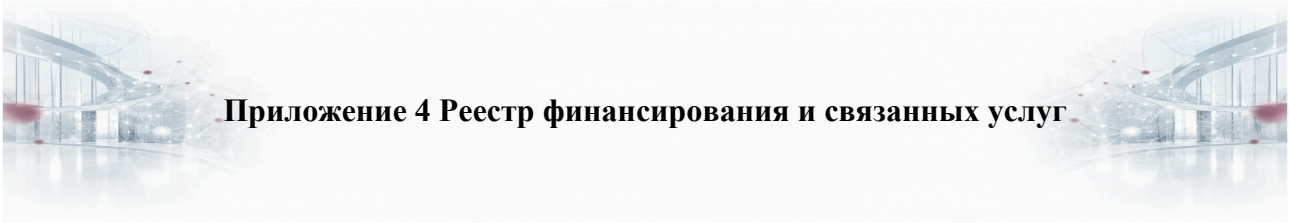
Порядок применения и ограничения

Решение о конфликте принимает лицо или комитет, не участвующий в исследовании. Оно может содержать полное участие, участие с раскрытием, запрет работы над определёнными частями или отвод. Самопроверка эксперта недостаточна. Публично отражается процедура управления конфликтом, а не полный закрытый текст декларации.

Поле/контроль	Правило заполнения или рекомендуемое значение
ID проекта	Связь с конкретным заданием
Финансовая связь	Договор, грант, активы: релевантность и период
Профессиональная роль	Разработка, внедрение, консультация
Личные и иные связи	Только способные повлиять на вывод
Решение	Нет конфликта / раскрытие / ограничение / отвод
Проверяющий	Независимое от эксперта лицо

Проверка на практике

Пример оценки: технический эксперт был сотрудником поставщика проверяемой модели. Он может описать принципы модели и провести консультацию, однако решение о достоверности её результатов передаётся независимому reviewer. Если участие этого эксперта несущественно для итогового вывода, достаточно ограниченной роли. Если без него невозможно получить данные, исследование публикуется как имеющее ограниченную независимость или откладывается. Решение с датой сохраняется в реестре конфликта и должно быть доступно для внутреннего аудита.



Приложение 4 Реестр финансирования и связанных услуг

Назначение

Финансовый реестр необходим для оценки риска зависимости от одного источника и скрытого влияния на повестку. Он включает денежное финансирование, значительные услуги натурой, предоставление лицензий и исследовательских данных. Нельзя скрывать существенную поддержку за категорией «партнёрская помощь», если она обеспечивает критическую инфраструктуру мониторинга.

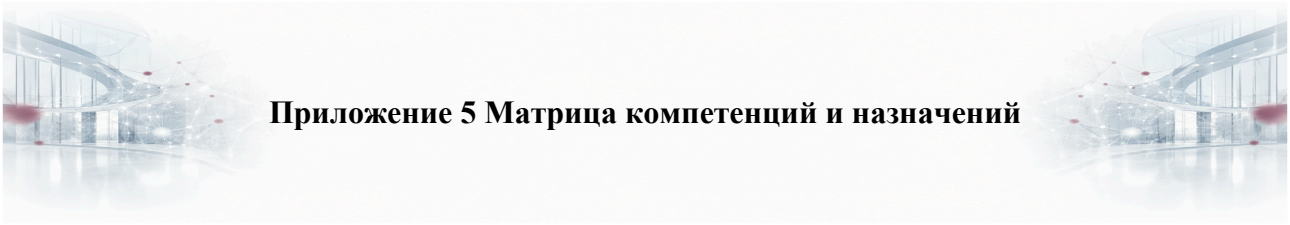
Порядок применения и ограничения

Ежегодный обзор должен показывать доли поступлений с ясным знаменателем; порог концентрации является внутренним сигналом управления, а не правовым запретом. Для закрытых договоров можно раскрывать категорию и долю, если точные условия подлежат допустимой защите. Совет регулярно оценивает право донора прекратить финансирование и зависимость от этого риска.

Поле/контроль	Правило заполнения или рекомендуемое значение
Источник	Наименование или правомерно агрегированная категория
Форма	Деньги / доступ / оборудование / часы эксперта
Назначение	Базовое, тематическое, заказное
Риск влияния	Договорные условия и концентрация
Редакционный контроль	Не допускается
Проверка	Годовой независимый обзор

Проверка на практике

Финансовые поступления группируются по категориям, а риски оцениваются не только в процентах доходов. Два договора одинакового размера могут иметь разные последствия: один не влияет на редакцию, другой содержит право согласования выводов. Для предоставленных бесплатно вычислительных ресурсов фиксируется возможность доступа дарителя к непубличным данным. Отчёт должен раскрывать метод оценки существенности in-kind поддержки и ситуацию, когда величину нельзя установить. Избыточная точность условных оценок не допускается.



Приложение 5 Матрица компетенций и назначений

Назначение

Для каждой проверки фиксируются необходимые компетенции. Правовой специалист не подменяет инженера, инженер не устанавливает юридическую ответственность без соответствующей квалификации, а редактор не решает вопрос допустимости сбора персональных данных исключительно стилистической оценкой. Модель назначения должна предотвращать концентрацию всех функций у одного сотрудника.

Порядок применения и ограничения

Матрица применяется до доступа к исходникам. Если нужной компетенции нет, задание сужается или привлекается внешний специалист. В заключении указывается фактическая область профессиональной проверки каждого участника, что особенно важно при использовании материала в судах или публичных консультациях.

Поле/контроль	Правило заполнения или рекомендуемое значение
Право	Применимая юрисдикция и сектор
Техническая оценка	Модель и тестовая методология
Статистика	Выборка, ошибки и доверительные интервалы
Защита данных	Основание и безопасность обработки
Редакция	Ясность и качество подтверждения
Внешнее review	При высокой общественной значимости

Проверка на практике

В матрице назначений для проверки автоматизированного санкционного скрининга участвуют специалист по соответствующему санкционному режиму, инженер идентификационных алгоритмов и эксперт по процедурам обжалования. Один юрист по общему праву не обеспечивает всех трёх квалификаций. Назначающий комитет проверяет конфликты до выдачи доступа. После анализа в отчёте перечисляются не должности ради престижности, а конкретные действия каждого reviewer: изучил норму, повторил тест, проверил данные или оценил способ публикации.



Приложение 6 Паспорт наблюдаемой AI-системы

Назначение

Карточка системного реестра отражает реально установленное применение, а не торговую марку. Она содержит информацию об операторе, целях, затронутой группе, типе автоматизированного вывода и наличии человеческого участия. При отсутствии документального подтверждения эксплуатации карточка получает предварительный статус и исключается из агрегированной статистики применяемых систем.

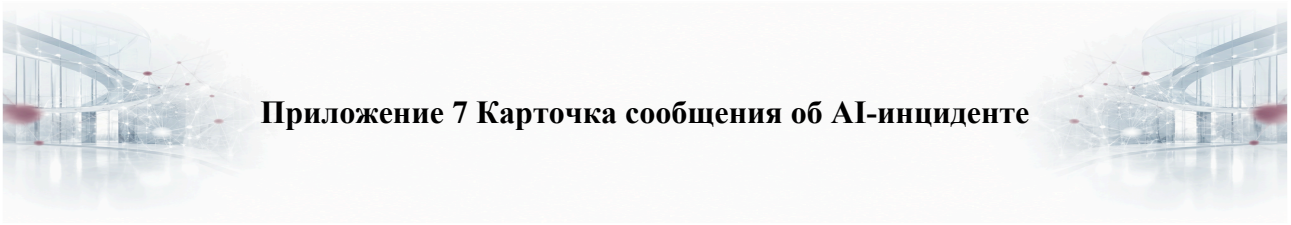
Порядок применения и ограничения

Каждое обновление получает номер версии. К карточке привязываются реестр рисков и события; изменение модели не удаляет предыдущие наблюдения. Данные о поставщике, национальных органах и применимом праве подтверждаются первичными источниками. Для закрытого продукта допустимо «версия не раскрыта», но не придуманное техническое описание.

Поле/контроль	Правило заполнения или рекомендуемое значение
Система/версия	Устойчивый идентификатор
Назначение	Решение, рекомендация, ранжирование, генерация
Деплойер	Фактический оператор использования
География	Где применяется, где данные
Состояние	Пилот / эксплуатация / прекращено / неизвестно
Доказательство	Документ и дата проверки
Правовая карта	Категория и дата применимости нормы

Проверка на практике

Если одно программное наименование используется пятью ведомствами с разными наборами данных и порогами, создаётся одна родительская запись и пять эксплуатационных карточек. Только так возможно различить дефект базовой технологии и ошибку локальной настройки. Дата закупки записывается отдельно от даты начала эксплуатации. В случае отсутствия независимого подтверждения внедрения карточка сохраняется как «проект анонсирован», не попадая в количественный показатель используемых AI-систем.



Приложение 7 Карточка сообщения об AI-инциденте

Назначение

Форма отделяет описание вреда от гипотезы о его причине. Заявитель может документировать решение и последствия, не зная архитектуры модели; отсутствие этих сведений не является основанием автоматически отказать в регистрации обращения. Однако

публичный статус AI-инцидента присваивается только после проверки роли ИИ и связи с актуальным вредом.

Порядок применения и ограничения

Обстоятельства дела хранятся в закрытом контуре до оценки безопасности публикации. Если речь идёт о нескольких потерпевших, создаются связанные эпизоды. Временная шкала записывается в нейтральной форме, с указанием источника каждого события. Проверка должна допускать конкурирующие причинные объяснения.

Поле/контроль	Правило заполнения или рекомендуемое значение
ID события	Уникальный и стабильный
Дата	Время события и время сообщения отдельно
Вред	Актуальный вред / возможный риск / неизвестно
Роль ИИ	Подтверждённая, вероятная или не установленная
Источник	Первичный документ или сообщение
Кому направлен вопрос	Организация и дата
Публичный статус	Непублично / предварительно / подтверждено

Проверка на практике

Условная регистрация: «клиент банка сообщает о длительной блокировке после автоматизированного сигнала». На входе устанавливается только факт обращения; нет доказательства незаконности блокировки или природы алгоритма. Аналитик запрашивает допустимые письма и документальные сведения, фиксирует сроки и проверяет наличие альтернативных причин. После анализа карточка может остаться сигналом либо перейти в подтверждённый инцидент, если установлены актуальный вред и причинная связь. Все изменения статуса остаются в журнале версий.

Приложение 8 Регистр AI-hazards и предварительных сигналов

Назначение

Реестр потенциальных опасностей необходим для раннего предупреждения без необоснованного утверждения о причинённом ущербе. Каждая запись описывает механизм потенциального вреда, необходимые условия, затронутые категории и уровень доказательств. Особое внимание уделяется необратимым последствиям, где ожидание реального инцидента противоречило бы разумной профилактике.

Порядок применения и ограничения

Переход из сигнала в hazard или incident оформляется как новая версия, а не редактирование без истории. Публичные числа hazards и incidents показываются раздельно. Предположение внешнего эксперта не становится подтверждённым фактом из-за его профессиональной известности.

Поле/контроль	Правило заполнения или рекомендуемое значение
Гипотеза	Проверяемый потенциальный механизм
Триггер	Когда риск способен реализоваться
Ущерб	Категория и возможная тяжесть
Подтверждение	Какие факты поддерживают вероятность

Поле/контроль	Правило заполнения или рекомендуемое значение
Меры	Что может предотвратить вред
Статус	Сигнал / hazard / incident

Проверка на практике

Для hazard не требуется ждать наступления реального вреда, но необходимо показать правдоподобную цепочку от системы к потенциальному последствию. Например, отсутствие механизма корректировки неверной идентификации в публичной услуге может создавать риск лишения помощи. Отдельно следует проверить, существует ли ручной резервный процесс. При обнаружении достаточных мер защиты уровень риска пересматривается. Такой пересмотр не должен стирать факт первоначальной гипотезы и её источников.

Приложение 9 Реестр рисков и контрольных мер

Назначение

Реестр рисков фиксирует не уровень репутационной опасности для разработчика, а сценарий возможного вреда для людей, организаций, инфраструктуры или окружающей среды. Карточка связывается с конкретной системой, контекстом и источником. Для сравнения в пределах одного проекта применяются внутренние уровни приоритета, но запрещено выдавать балл за универсальную правовую оценку.

Порядок применения и ограничения

При смене модели, появлении нового регуляторного решения или подтверждённой мере защиты риск переоценивается. Следует различать «мера заявлена» и «мера проверена». Остаточный риск остаётся открытым, если отсутствуют данные о реальной эффективности. Критические сценарии немедленного вреда эскалируются независимо от совокупного численного балла.

Поле/контроль	Правило заполнения или рекомендуемое значение
Сценарий	Обстоятельства → механизм → последствия
Тяжесть	Обратимость, охват, уязвимость
Уверенность	Уровень доказательств
Контроль	Заявлен / проверен / неизвестен
Остаток	Оценка после мер защиты
Пересмотр	Триггер, дата и ответственный

Проверка на практике

Риск-карточка должна содержать не только исходный уровень, но и подтверждение конкретной меры снижения. Если разработчик пишет «фильтр обновлён», это статус заявления. Подтверждением эффективности может служить повторная независимая проверка в том же контексте. Отдельно сохраняется дата изменения модели, чтобы сравнение до и после не приписывало результат другим обстоятельствам. Неопределённость остаточного риска явно указывается даже при публичном обещании поставщика.



Приложение 10 Словарь ключевых показателей

Назначение

Каждая метрика имеет отдельную паспортную запись, благодаря чему разные команды и страны не переопределяют показатели незаметно. Поля включают определение, числитель, знаменатель, период, источник, исключения, обработку дублей и качество данных. При отсутствии знаменателя метрика не называется частотой или распространённостью.

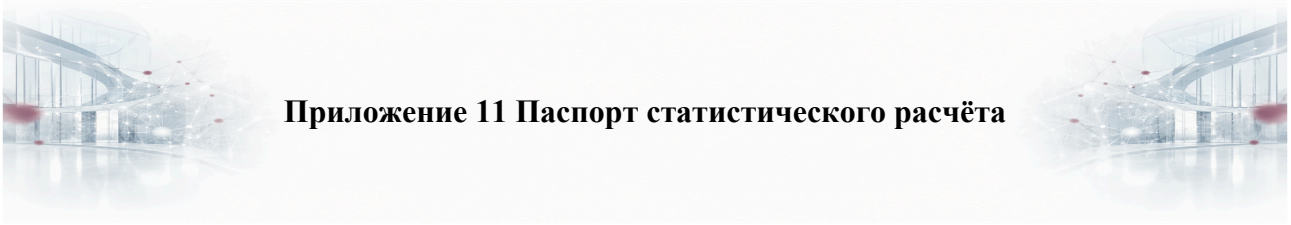
Порядок применения и ограничения

Версии словаря публикуются в открытом виде, а при изменении определения исторический ряд либо пересчитывается, либо разрывается на две несопоставимые серии. Особые предупреждения нужны для количества обращений: оно отражает и реальный вред, и изменение доверия к каналу сообщения.

Поле/контроль	Правило заполнения или рекомендуемое значение
M-01	Подтверждённые инциденты: абсолютное число
M-02	Hazards: отдельная серия
M-03	Доля карточек с первичными источниками
M-04	Время от сообщения до предварительного решения
M-05	Доля ответов организаций на запрос
M-06	Доля опубликованных исправлений
Обязательное поле	Наличие или отсутствие проверяемого знаменателя

Проверка на практике

Для показателя времени рассмотрения обращения необходимо определить момент начала отсчёта, например регистрацию полного обращения, и момент завершения, например мотивированное решение по проверке. Если часть обращений ещё открыта, среднее время среди закрытых случаев создаст смещение. Поэтому dashboard дополнительно показывает число открытых дел и их возраст. Для доли ответов организаций учитывается, был ли направлен конкретный проверяемый запрос и истёк ли разумный срок. Иначе показатель наказал бы адресатов, которым запрос не направляли.



Приложение 11 Паспорт статистического расчёта

Назначение

Форма применяется всякий раз, когда отчёт использует доли, сравнения периодов или интервалы. Пример является условным: 12 проверенных ошибок на 300 проанализированных решений дают 4%, но это частота только внутри проанализированной выборки, а не во всей организации. Если решения были выбраны по жалобам, результат не может быть распространён на общий поток клиентов.

Порядок применения и ограничения

Доверительные интервалы требуют установленной модели отбора. Для целенаправленно собранной выборки интервал биномиальной случайной выборки может вводить читателя в заблуждение. Отчёт должен прямо указать источник смещения и отказ от генерализации.

Поле/контроль	Правило заполнения или рекомендуемое значение
Единица	Решение / пользователь / система / жалоба
Числитель	Наблюдений с заданным признаком
Знаменатель	Все подходящие объекты в выборке
Отбор	Случайный / сплошной / по обращениям
Дубли	Правила объединения
Результат	Оценка только для подходящей совокупности
Ограничение	Почему нельзя автоматически обобщать

Проверка на практике

В условном числовом примере 12 из 300 равны 4%, однако текст обязан раскрывать отбор этих 300. Если исследовались только поступившие жалобы, нельзя писать, что у организации ошибочны 4% всех решений. При сравнении с другим периодом проверяется изменение определения ошибки и доли неполных записей. При недостатке данных вывод формулируется «4% обследованной выборки», без экстраполяции на всех клиентов и без искусственной точности после запятой.



Приложение 12 Классификатор качества источников

Назначение

Классификатор не заменяет анализа содержания. Один и тот же документ может быть сильным доказательством существования решения, но слабым подтверждением технических причин. Для каждого утверждения отдельно фиксируется, какой факт источник действительно устанавливает. Газетная статья с цитатой официального решения не становится самим решением.

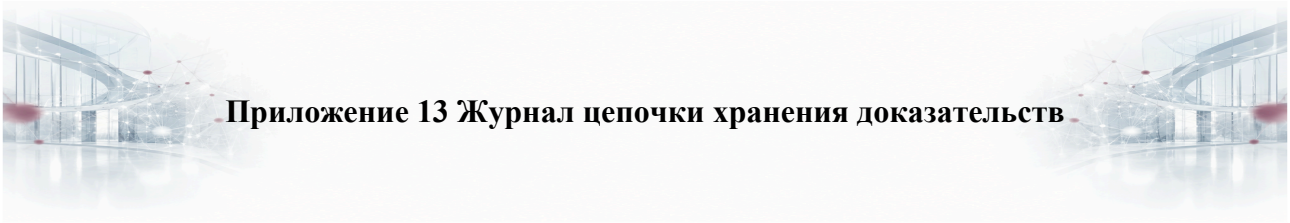
Порядок применения и ограничения

Если источник противоречит другому, конфликт не разрешается механическим повышением ранга. Аналитик ищет оригиналы, даты и процессуальный статус, а при невозможности получения документов оставляет вывод ограниченным. В публичном тексте указывается характер подтверждения.

Поле/контроль	Правило заполнения или рекомендуемое значение
A	Официальный первичный документ с проверяемым происхождением
B	Прямой материал участника с подтверждающими приложениями
C	Независимая публикация с проверяемыми ссылками
D	Неидентифицированное или неподтверждённое сообщение
Дополнение	Происхождение и содержание оценивать отдельно
Итог	Качество для конкретного тезиса, не автора в целом

Проверка на практике

В конфликте источников официальный акт может подтверждать, что решение существовало, но не обязательно все фактические утверждения стороны. Вторичная аналитическая статья может верно объяснять контекст, но должна вести к первичному источнику. Для спорного тезиса reviewer оценивает содержательную относимость каждого документа. Если оригинал не найден, ссылка на цитирование не объявляется доказательством первичного уровня. Такой способ позволяет честно изложить доступное знание без необоснованного ранжирования людей по доверию.



Приложение 13 Журнал цепочки хранения доказательств

Назначение

Журнал позволяет восстановить поступление, преобразование и использование электронного файла. Он не создаёт презумпции подлинности по одному наличию хэша: необходима проверка происхождения и условий получения. Для закрытых материалов дополнительно фиксируются правовое основание, разрешение на доступ и способ защиты персональных сведений.

Порядок применения и ограничения

При производной обработке, например удалении личных данных, сохраняются связь с исходником и сведения об операции. В публичную версию выносится только допустимый идентификатор. Если юридическое хранение исходника невозможно, фиксируется основание удаления и характеристики доказательственной потери.

Поле/контроль	Правило заполнения или рекомендуемое значение
EVID-ID	Номер материала
Время	UTC и местное время при необходимости
Получено от	Источник, канал и основание
Исходник	Формат, контрольная сумма
Операция	Копирование, перевод, маскирование, анализ
Доступ	Лицо, время, цель
Удаление	Дата и законное основание

Проверка на практике

При копировании файла сохраняются исходные байты, время получения и новый хэш рабочей копии. Перевод отделяется от оригинала и сопровождается обозначением переводчика или использованного средства; при существенном юридическом термине необходима профильная проверка. Маскирование персональных данных должно быть необратимым в опубликованном файле, а не поверхностной закраской текста. Если лог передачи доказательств неполон, заключение прямо описывает этот пробел и его возможное влияние на результат.

Приложение 14 Протокол независимой проверки карточки

Назначение

Второй исследователь должен проверить роль ИИ, категорию события, первичные документы, соответствие временной линии и ограничения вывода. Нельзя ограничиваться проверкой орфографии или существования указанных URL. Для технических заявлений дополнительно описывается возможность воспроизвести тест и модель/версию, к которой он относился.

Порядок применения и ограничения

Рецензент заполняет мотивированное решение: подтверждено, подтверждено с оговорками, требуется доработка, недостаточно данных, отклонено. При несогласии сохраняется запись обеих позиций и применённый стандарт доказательств. Публикация без разрешения существенного противоречия должна иметь явное предупреждение.

Поле/контроль	Правило заполнения или рекомендуемое значение
Факт	Что доказано и чем
Альтернатива	Какое объяснение не исключено
Источник	Проверен ли оригинал
Право	Нет ли юридически ошибочного вывода
Данные	Законность и безопасность раскрытия
Решение	Подтверждение / доработка / отказ

Проверка на практике

Рецензент задаёт обязательный вопрос: может ли другой специалист по тем же исходным документам прийти к иному разумному выводу? Если да, следует описать альтернативу и степень неопределённости. Например, решение государственного органа может быть результатом неверных данных, а не дефекта модели. Даже если использование ИИ известно, необходимо доказать его значимую роль. В протоколе review такие альтернативы фиксируются до утверждения заголовка публикации.

Приложение 15 Процедура защиты чувствительного заявителя

Назначение

Форма используется при риске преследования, разглашения миграционного статуса, банковских обстоятельств, состояния здоровья или семейных деталей. Перед публикацией специалист отдельно оценивает прямые идентификаторы и сочетание косвенных признаков. Имя и место могут быть скрыты, но редкая хронология сохраняет риск узнавания.

Порядок применения и ограничения

Решение о публикации не может опираться только на интерес СМИ. При высоком риске разрешён агрегированный обзор без индивидуальной истории. Если информация поступила через адвоката или доверенное лицо, необходимо проверить пределы полномочий на распространение, не предполагая, что право представлять в деле равно согласию на публичное раскрытие.

Поле/контроль	Правило заполнения или рекомендуемое значение
Риск	Физический, социальный, миграционный, экономический
Идентификаторы	Прямые и косвенные
Минимизация	Какие детали действительно нужны
Согласование	Отдельно от законности обработки
Альтернатива	Агрегат или композитный сценарий
Итог	Публиковать / ограничить / не публиковать

Проверка на практике

Для человека, ищущего убежище, сочетание редкой даты задержания, места и исхода процедуры может однозначно идентифицировать личность даже без имени. Редактор проверяет возможность поиска по такой комбинации в доступных базах и СМИ. При высоком риске изменённые детали композитного сценария должны быть обозначены как условные. Если детализация необходима для юридической аргументации, материал может оставаться в закрытой экспертной версии без публичного распространения.

Приложение 16 Стандарт запроса о комментарии и праве на ответ

Назначение

Запрос направляется после проверки значимых фактических оснований и до публикации существенного негативного вывода, если нет оправданного исключения. Он излагает конкретные обстоятельства, вопросы и документы, которые могут изменить оценку. Недопустимо отправлять эмоциональное обвинение с требованием немедленно «опровергнуть» без указания рассматриваемого тезиса.

Порядок применения и ограничения

Срок ответа фиксируется и соразмеряется сложности. Если ответ поступил, аналитик отдельно указывает, какие сведения приняты и какие остались неподтверждёнными. При отсутствии ответа используется нейтральная формулировка, не равнозначная признанию. Личная переписка сотрудников не публикуется, если она не необходима и не допустима по праву.

Поле/контроль	Правило заполнения или рекомендуемое значение
Адресат	Ответственное лицо либо организация
Тезис	Конкретное и проверяемое утверждение
Запрос	Документы, причины, технический контекст
Срок	Разумный, заранее сообщённый
Конфиденциальность	Режим ответа
Результат	Получен / частично / не получен
Учёт	Где это отразилось в отчёте

Проверка на практике

Письмо с запросом комментария должно содержать вопрос, на который реально можно ответить, и понятную дату закрытия материала. Например: «Подтверждаете ли вы использование данного классификатора в указанном процессе и существует ли документированный человеческий пересмотр?» Нельзя подразумевать виновность из отказа отвечать. При получении ответа необходимо сравнить его с первичными документами и отметить разницу между подтверждением факта, возражением против интерпретации и обещанием изменить процедуру.

Приложение 17 Порядок исправления, отзыва и замены отчёта

Назначение

Процедура отвечает на типовую проблему: исследовательская организация публикует важный вывод, затем обнаруживает ошибку в дате, идентификации лица или юридическом статусе документа. При несущественной ошибке достаточно публичного erratum; при материальной — новой версии с заметным уведомлением; при риске прямого вреда — срочной блокировки неверного фрагмента с мотивированным журналом решений.

Порядок применения и ограничения

Окончательное решение о корректировке принимает редакционный орган, не связанный с первоначальным спором. История сохраняется настолько, насколько это совместимо с правом на защиту данных. Пользователям материалов должны быть доступны актуальная версия и указание, где существенный вывод изменился.

Поле/контроль	Правило заполнения или рекомендуемое значение
Тип ошибки	Техническая / фактическая / правовая / безопасность
Последствия	Есть ли прямой ущерб от сохранения
Действие	Erratum / обновление / отзыв
Рецензия	Независимая проверка при существенном споре
Версия	Стабильный номер и дата
Уведомление	Кому направить информацию об исправлении

Проверка на практике

Для опечатки в малозначимой дате подходит erratum. Для ошибочного указания на причастность конкретного человека требуется существенное исправление, возможно с уведомлением адресатов, уже получивших отчёт. Для обновления закона после даты исследования выпускается новая актуализированная версия, а не переписывается историческая оценка. При публикации исправления сохраняется достаточная информация о старой ошибке, насколько это совместимо с защитой персональных данных.

Приложение 18 Матрица применимости права по юрисдикции

Назначение

Эта таблица предназначена для предварительной карты, а не заменяет правового заключения. Отдельно фиксируются статус акта, его обязательность, применимые статьи, дата начала действия для данного класса системы и компетентный орган. Особое значение имеет различие подписания и вступления международного договора в силу, а также добровольных стандартов и правовых обязательств.

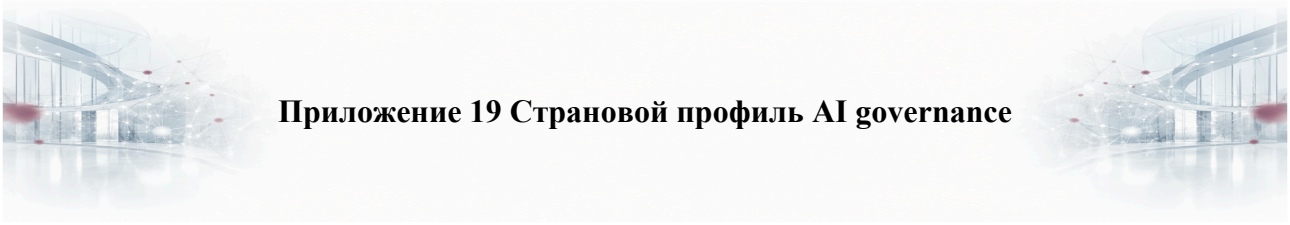
Порядок применения и ограничения

Проверка проводится на дату рассматриваемого события и повторно перед публикацией. Для AI Act требуется актуальная консолидированная редакция; для Framework Convention — депозитарный статус; для GDPR — роль контролёра и характер обработки. При недостатке сведений нельзя присваивать статус «нарушение установлено».

Поле/контроль	Правило заполнения или рекомендуемое значение
Юрисдикция	Страна и применимый уровень права
Инструмент	Закон / договор / стандарт / руководство
Статус	Действует / переходный / не обязывает
Роль	Поставщик, оператор, контролёр и др.
Доказательство	Официальный источник и дата
Компетенция	Кто вправе применять санкции

Проверка на практике

Условная таблица права может содержать колонку «акт действует» и отдельную «обязанность применима к субъекту». Для AI Act дата и категория системы иногда различаются; для международной конвенции подпись не равна ратификации; для стандарта ISO сама публикация не создаёт общей юридической обязанности. В случае сомнения итоговое поле содержит «требуется заключение по применимости». Такая отметка защищает от эффектных, но неверных фраз «нарушает международный стандарт» без объяснения правовой силы.



Приложение 19 Страновой профиль AI governance

Назначение

Профиль объединяет правовые сведения, институциональную инфраструктуру, способы защиты пострадавших и качество данных. Он не является международным рейтингом государств и не предполагает, что отсутствие публичных нарушений показывает успешный контроль. Полнота жалоб может быть выше именно там, где процедуры защиты более доступны.

Порядок применения и ограничения

Источники страны указываются с датой и языком. Все сравниваемые категории снабжаются меткой сопоставимости. Заключения о реальном исполнении права требуют самостоятельных доказательств: опубликованных проверок, судебных решений и измеримых результатов, а не текста закона как такового.

Поле/контроль	Правило заполнения или рекомендуемое значение
Право	Обязательства и фактическая дата
Органы	Полномочия и публичные каналы
Практика	Проверенные случаи и решения
Защита	Жалобы, пересмотр, судебные средства
Данные	Периоды, охват, неопределённость
Сравнимость	Что нельзя напрямую сопоставить

Проверка на практике

Страна с развитой системой публичных жалоб может иметь больше открытых кейсов, чем страна с ограниченными средствами защиты. Сравнение только чисел жалоб перевернуло бы возможный вывод. Поэтому профиль содержит оценку доступности каналов, языкового покрытия и объёма публичных решений. Если данные по этим параметрам недоступны, Observatory не строит единый рейтинг качества управления ИИ, а публикует отдельные проверяемые характеристики.

Приложение 20 Паспорт методики международного сопоставления

Назначение

Сравнительный доклад обязан объяснять, почему выбранные страны и классы объектов являются сопоставимыми. Например, государство с открытым реестром закупок имеет больше наблюдаемых внедрений, чем государство без такого реестра; прямой рейтинг числа AI-систем был бы искажён. Выборка стран фиксируется до получения результатов, а исключения объясняются.

Порядок применения и ограничения

При разнотипных источниках сравниваются отдельные домены, а не единый итоговый балл. Для динамики используются одинаковые определения по годам; при изменении методики публикуется новое основание ряда. Межъязыковая проверка обеспечивает корректный перевод правовых терминов, не стирая особенности национального права.

Поле/контроль	Правило заполнения или рекомендуемое значение
Сравниваемый объект	Система, сектор или процедура
Страны	Критерий отбора
Период	Единая дата среза
Определения	Перевод и совпадение смысла
Источники	Классы и покрытие
Вывод	Сопоставимо / частично / несопоставимо

Проверка на практике

Для международного сравнения санкционного скрининга важна единая единица наблюдения. «Предупреждение», «ограниченная операция», «проверенный клиент» и «оспоренное решение» — разные сущности. Даже одинаковая доля отменённых ограничений не означает одинаковой технической точности, если отличаются нормы, клиентские базы и политика рассмотрения. Протокол требует отдельной таблицы определений для каждой юрисдикции и честного заключения о несопоставимости при отсутствии таких данных.

Приложение 21 Методика отраслевого скрининга AML и санкций

Назначение

Это исследовательская, а не банковская операционная инструкция. Для качества анализа требуется отделять сигнал совпадения от окончательного решения, юридическую обязанность от внутреннего аппетита к риску, идентификационный спор от санкционного статуса. Публикация не должна содержать конкретные настройки контроля, позволяющие недобросовестному лицу обходить мониторинг.

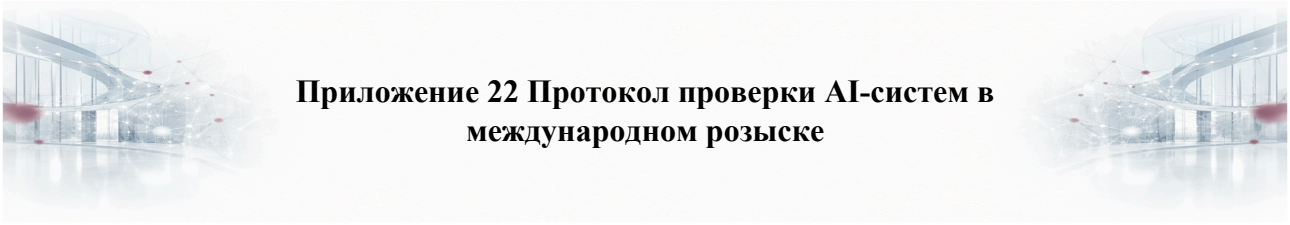
Порядок применения и ограничения

В тестовом наборе следует учитывать сложности транслитерации и повторяющихся имён. Результаты маркируются по природе ground truth; если независимой достоверной разметки нет, технические error rates не выводятся. Отдельно оцениваются скорость пересмотра и качество объяснения клиенту в предусмотренных правом пределах.

Поле/контроль	Правило заполнения или рекомендуемое значение
Объект	Алгоритмический сигнал
Правовая роль	Банк / провайдер / клиент / регулятор
Разметка	Источник истинного статуса
Смещение	Имена, языки, группы
Пересмотр	Результат, срок, основание
Публикация	Без чувствительных порогов контроля

Проверка на практике

Тест на условном наборе транслитераций включает похожие имена, различия дат рождения, перестановки частей фамилии и варианты письма. Для реальных санкционных режимов итоговое решение требует юридической проверки, поэтому лабораторный результат не объявляется основанием разморозки средств. Команда отдельно документирует, какие типы ошибок обнаружены, и не раскрывает эксплуатационные пороги, способные облегчить обход контроля. В публичной части описывается процесс человеческой проверки и право на доступные процедуры исправления.



Приложение 22 Протокол проверки AI-систем в международном розыске

Назначение

Протокол направлен на оценку публично доступных процедур, а не на подтверждение статуса конкретного человека в закрытых системах. Аналитик раздельно описывает национальное решение, международный канал обмена данными, техническое совпадение и последующее судебное рассмотрение. Ни один этап не выводится автоматически из другого.

Отдельное условие допустимости: никакая исследовательская карточка не должна превращаться в доступную по поиску базу обвинений или сведений о судимости. Входящие материалы проверяются на подпадание под статью 10 GDPR и применимые национальные правила до их помещения в централизованное хранилище. Публикация содержания уведомления, факта задержания либо решения об экстрадиции не допускает автоматического вывода, что в событии участвовал ИИ или что указанное лицо действительно совершило вменяемое деяние. [13]

Порядок применения и ограничения

Использование персональных дел для публичной аналитики требует отдельной оценки права и безопасности. Реальные имена, изображения и биометрия по умолчанию исключаются из открытой версии. Если нет технического подтверждения применения алгоритма, вывод касается качества процесса и вероятного риска, а не конкретного AI-дефекта.

Поле/контроль	Правило заполнения или рекомендуемое значение
Канал	Какая процедура фактически подтверждена
Документ	Орган, дата, правовой эффект
AI-компонент	Доказанная функция или неизвестно
Человеческий контроль	Кем и когда проверялось
Процедура защиты	Наличие доступа к исправлению
Публикация	Деперсонализация и ограничения

Проверка на практике

Если судебный акт подтвердил освобождение человека после задержания, он устанавливает именно процессуальный факт и свои собственные выводы, а не автоматически устройство информационной системы, приведшей к задержанию. Исследователь может оценить качество цепочки контроля, наличие письменного основания и доступ к предусмотренным процедурам исправления. Он не может по открытой публикации подтвердить состояние

непубличной базы INTERPOL. Для такого тезиса нужны компетентные официальные средства доступа и отдельная юридическая оценка.

Приложение 23 Матрица человеческого контроля и обжалования

Назначение

Матрица помогает отличать формальное наличие человека от реальной возможности изменить алгоритмический результат. Нужны сведения о полномочиях сотрудника, доступе к исходным данным, времени на проверку и независимости от автоматизированной рекомендации. Простая подпись должностного лица не доказывает содержательного пересмотра.

Порядок применения и ограничения

Для пострадавшего важны уведомление, понятная процедура запроса и наличие способа изменить ошибочный результат. Отдельно оценивается, способен ли орган предоставить достаточное объяснение с учётом законных ограничений. Исследователь не подменяет решение компетентной инстанции.

Поле/контроль	Правило заполнения или рекомендуемое значение
Сигнал	Когда появляется автоматизированный вывод
Человек	Право изменить итог
Данные	Доступ к сведениям для независимой проверки
Уведомление	Что узнаёт затронутое лицо
Жалоба	Фактический порядок и срок
Итог	Есть ли подтверждение пересмотра

Проверка на практике

Для публичной услуги особенно важна возможность «разморозить» автоматизированный результат, когда появились новые документы. Необходимо выяснить, кто имеет право изменить итог, сохраняется ли запись о причинах исправления и могут ли заявители узнать о процедуре. Формальная галочка «решение просмотрел сотрудник» не доказывает независимой оценки. Наличие полноценного пересмотра подтверждается документами, полномочиями и, по возможности, примерами корректировки ошибочных решений.

Приложение 24 Форма отчёта о внешней рецензии

Назначение

Рецензия должна оценивать доказательственную сторону, а не репутационную удобность текста. В ней указываются компетенция эксперта, конфликты интересов, границы рассмотренных материалов, методы проверки и конкретные критические замечания.

Рецензент не обязан объявлять полный текст «достоверным» только потому, что его попросили прочитать отдельные разделы.

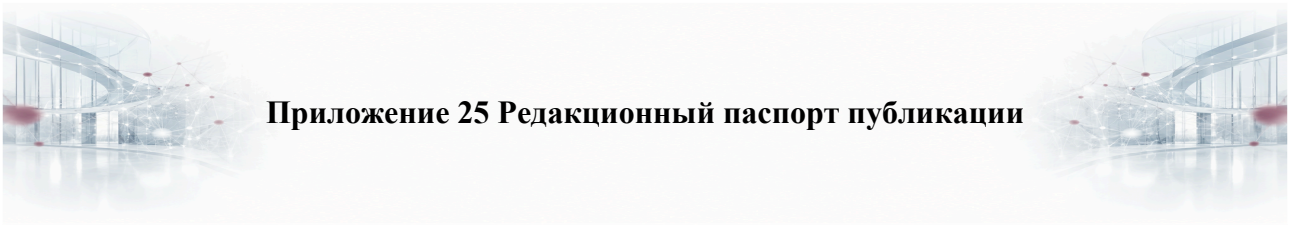
Порядок применения и ограничения

Редакция связывает каждое существенное замечание с изменением версии либо мотивированным отклонением. В публичном документе может быть указан режим review — внутреннее методическое, внешнее научное, правовое, статистическое — без приписывания более высокой процедуры, чем реально проводилась.

Поле/контроль	Правило заполнения или рекомендуемое значение
Рецензент	Компетенция и независимость
Предмет	Полный отчёт или ограниченный раздел
Замечание	Точная цитата или номер тезиса
Серьёзность	Критическая / существенная / техническая
Решение автора	Принято / отклонено с мотивировкой
Итог	Достаточно для выпуска или нет

Проверка на практике

В отчёте внешней рецензии критическое замечание должно относиться к конкретному утверждению: например, «не установлен знаменатель процента отказов» или «вместо правового текста использован вторичный пересказ». Общая фраза «доклад качественный» мало пригодна для проверки. Автор документирует исправление, а reviewer при необходимости подтверждает его. В публичной версии указан вид review, чтобы читатель не путал техническую вычитку с независимой научной рецензией.



Приложение 25 Редакционный паспорт публикации

Назначение

Паспорт закрепляется за каждой публичной работой и позволяет последующему читателю оценить пределы исследовательской ответственности. Он не должен подменяться одной фразой «материалы проверены». Указание периода сбора, методов отбора и уровней источников особенно важно для динамических AI-систем.

Порядок применения и ограничения

При переиздании паспорт обновляется вместе с номером версии. Ссылки на законодательство проверяются на дату выпуска, а материалы по коммерческим системам — после значимых обновлений. Публичная страница должна вести к актуальной версии и журналу исправлений.

Поле/контроль	Правило заполнения или рекомендуемое значение
Название	Устойчивый идентификатор работы
Дата среза	Последний день включённой информации
Авторы	Ответственность по разделам
Метод	Ссылка на версию протокола
Источники	Первичные и вторичные отдельно
Ограничения	Недоступные данные и смещения
Версия	Дата выпуска и исправления

Проверка на практике

Редакционный паспорт не должен быть декоративным. Если после публикации выяснилось, что одна категория источников стала недоступной, новая версия указывает изменение полноты данных. Устаревший закон, отмеченный без даты действия, подлежит уточнению до распространения. Для динамических моделей особенно важно записывать версию и дату тестирования; оценка прежней версии не распространяется автоматически на обновлённый продукт. История паспорта помогает использовать материал в дальнейших процедурах без искажения его первоначальных оснований.

Приложение 26 Шаблон обращения к регулятору или публичному органу

Назначение

Обращение от независимого исследователя должно быть точным и не создавать ложное впечатление, что организация действует от имени международного органа. В нём указываются проверенные факты, источник информации, сфера компетенции адресата и конкретная просьба: предоставить публичные сведения, разъяснить применимую процедуру или рассмотреть проблему в пределах полномочий.

Порядок применения и ограничения

Персональные материалы направляются только при наличии законного основания и достаточной необходимости. Подтверждение получения сохраняется в реестре контактов, но не выдаётся за согласие органа с содержанием исследования. Если адресат не обладает компетенцией, обращение корректируется, а не распространяется в массовой рассылке.

Поле/контроль	Правило заполнения или рекомендуемое значение
Отправитель	Правильно указанный юридический статус
Предмет	Краткое описание наблюдения
Факты	Нумерованные документы и даты
Компетенция	Почему выбран адресат
Просьба	Конкретное законное действие
Приложения	Минимальный достаточный объём
Подтверждение	Регистрационный номер или ответ

Проверка на практике

В обращении к международному органу нельзя писать, что Observatory «уполномочен проверить государство», если такой мандат не предоставлен. Корректная формулировка — «просим рассмотреть публично документированные сведения в пределах вашей компетенции». Отправка и получение фиксируются отдельно от результата рассмотрения. Подтверждение электронной доставки не означает поддержки тезиса адресатом. Если речь идёт о персональных данных, объём приложения сокращается до законно необходимого минимума.

Приложение 27 Протокол срочной аналитической эскалации

Назначение

Срочный протокол предназначен для случаев вероятного необратимого вреда, когда ожидание очередного квартального отчёта неразумно. Он не снижает стандарт честного описания доказательств, но допускает узкий предупредительный запрос: например, просьбу сохранить логи и представить контакт для конфиденциального обмена. Публичное обвинение без подтверждения остаётся недопустимым.

Порядок применения и ограничения

Решение об эскалации принимает специально назначенная пара специалистов, включая юридическую проверку риска раскрытия. После выхода из срочного режима проводится ретроспективный анализ, не был ли механизм использован для обхода нормального процесса review.

Поле/контроль	Правило заполнения или рекомендуемое значение
Триггер	Вероятный и тяжёлый вред
Доказательства	Что подтверждено на данный момент
Ограничение	Какой важной информации не хватает
Действие	Сохранить / запросить / направить компетентному органу
Защита	Минимум раскрываемых сведений
Контроль	Обязательный послесобытийный review

Проверка на практике

При риске немедленного вреда возможно срочно направить просьбу сохранить логи до завершения подробного исследования. Но сообщение должно содержать точные пределы известных фактов. Комитет не обязан заявлять «преступление доказано», чтобы обратить внимание на возможное уничтожение доказательств. После срочного этапа независимый эксперт оценивает, была ли процедура действительно обоснованной, и оформляет уроки для будущих случаев. Публичная история срочного обращения не должна раскрывать личность защищаемого человека.

Приложение 28 Правила визуализации публичного dashboard

Назначение

Публичный dashboard должен показывать проверенные данные и не создавать видимость точности там, где исследователь не знает знаменателя. Все графики имеют период, источник, определение и ссылку на метод. Число событий из СМИ не называется уровнем распространённости вреда. Для малых ячеек и персональных событий применяются правила агрегации и подавления чувствительных подробностей.

Порядок применения и ограничения

При изменении классификации история данных обновляется с заметным указанием разрыва ряда. Любые рейтинги должны проходить дополнительный review. Пользователь может скачать словарь показателей и объяснение того, почему некоторые сравнения не проводятся.

Поле/контроль	Правило заполнения или рекомендуемое значение
Метрика	Название и точное определение
Период	Начало и конец
Охват	Страны, языки, источники
Знаменатель	Есть или отсутствует
Качество	Доля подтверждённых записей
Защита	Правило малых выборок
Версия	Изменения классификатора

Проверка на практике

На графике числа инцидентов следует отдельно показать изменения покрытия источников. Если в первом году анализировались два языка, а во втором десять, рост числа зарегистрированных сообщений нельзя автоматически считать ростом вреда. Когда определение события меняется, данные пересчитываются или отмечается разрыв ряда. Для небольших стран и редких чувствительных событий статистическая таблица может привести к идентификации человека; в таком случае мелкие категории объединяются или скрываются.

Приложение 29 План пилота на двенадцать месяцев

Назначение

Предлагаемая последовательность условна и требует одобрения руководством. Месяцы 1–2 отводятся на устав и аудит правовых оснований; 3–4 — на инфраструктуру и обучение; 5–7 — на закрытую тестовую выборку; 8–9 — на независимое рецензирование; 10–11 — на выпуск двух тематических результатов; 12 — на проверку качества и решение о масштабировании.

Порядок применения и ограничения

Ключевое условие продолжения — не число опубликованных громких кейсов, а работоспособность процедуры исправления, отсутствие необработанных конфликтов и безопасность данных. При неудаче проект сокращается и повторно тестируется. Следует прямо указать, что расписание является проектным предложением и не подтверждает уже утверждённой программы мероприятий.

Поле/контроль	Правило заполнения или рекомендуемое значение
M1–M2	Право, устав, конфликты
M3–M4	Инфраструктура и словарь
M5–M7	Тестовый непубличный сбор
M8–M9	Методический review
M10–M11	Публикации с ограничениями
M12	Внешняя оценка и решение

Проверка на практике

Запланированные двенадцать месяцев — не подтверждённое расписание ARGA. До запуска необходимо проверить правовую форму, согласовать ресурсы и реальный штат независимой проверки. На непубличной выборке тестируются не только алгоритмы классификации, но и обработка жалоб, защищённый доступ и способность опровергнуть собственную гипотезу. Если один из критических элементов не функционирует, масштабирование откладывается. Вывод пилота может быть негативным — это приемлемый исследовательский результат.

Приложение 30 Сценарий: ложное санкционное совпадение

Назначение

Условный пример: банк приостанавливает обслуживание клиента после совпадения транслитерированного имени с записью перечня. Клиент предоставляет документы, подтверждающие отличия; решение пересматривается после проверки сотрудником. Само событие не доказывает ошибочности применённого санкционного закона. Observatory исследует, каким образом алгоритмическая подсказка перешла в ограничение и насколько был доступен пересмотр.

Порядок применения и ограничения

Документы кейса включают даты уведомлений, допустимые объяснения банка, юридические основания, результат пересмотра и процессуальную хронологию. Для оценки технической FPR не хватает множества истинно отрицательных случаев; поэтому доклад честно ограничивается исследованием процедуры и случая ошибочной идентификации.

Поле/контроль	Правило заполнения или рекомендуемое значение
Факт	Ограничение обслуживания
Гипотеза	Совпадение по имени связано с автоматизацией
Нужный документ	Основание решения и результат пересмотра
Запрет вывода	Не называть один случай общей FPR
Возможная рекомендация	Разделить сигнал и окончательное решение

Проверка на практике

В кейсе санкционного совпадения ключевой момент — переход от автоматического предупреждения к ограничению обслуживания. Нужно документировать, какой субъект принял решение и какой механизм пересмотра был доступен. Ошибка идентификации может быть подтверждена последующим решением, но это не означает, что весь санкционный контроль организации ненадёжен. Для технического процента ложных совпадений требуется проверенная совокупность истинно отрицательных случаев, которой индивидуальное дело не предоставляет.

Приложение 31 Сценарий: риск ошибочной идентификации на границе

Назначение

Условный пример: лицо задерживают после совпадения данных с международным уведомлением, затем освобождают по решению компетентного органа. Публичные документы подтверждают задержание и освобождение, но не раскрывают, была ли первопричиной ошибка алгоритма, неточность исходной записи или иная процедура. Observatory не вправе объявлять «сбой интерполовского ИИ» без технического основания.

Порядок применения и ограничения

Исследование ограничивается доступными фактами, правом на исправление записи и состоянием национального процесса. Имя и маршрут лица исключаются из публичной версии. В итоговом выводе допускается описание процессуального риска при автоматическом использовании совпадений без должного человеческого контроля.

Поле/контроль	Правило заполнения или рекомендуемое значение
Установлено	Процессуальные факты по документам
Не установлено	Конкретный AI-механизм совпадения
Риск	Недостаточный человеческий пересмотр
Защита	Консультация юриста и предусмотренные процедуры
Публикация	Полностью обезличенная модель

Проверка на практике

В кейсе международного розыска одно судебное освобождение не доказывает отсутствия записей в любых международных базах и не устанавливает AI-причину задержания. Observatory фиксирует процессуальную хронологию и проверяет, какой документ способен подтвердить каждое звено. Отсутствующие технические сведения отражаются явно. Публикация предназначена для анализа гарантий исправления, а не для раскрытия статуса конкретного заявителя, особенно если ему угрожает преследование.

Приложение 32 Сценарий: социальная выплата и непрозрачный скоринг

Назначение

Условный пример: заявителю отказано в выплате, а уведомление не объясняет, какие сведения привели к выводу. Выясняется, что сотрудник использовал цифровую рекомендацию. Это не позволяет автоматически утверждать незаконную автоматизацию: нужно проверить компетенцию, национальное право и возможность пересмотра. Для пострадавшего первоочередна доступность эффективной жалобы и исправления недостоверных данных.

Порядок применения и ограничения

Observatory готовит карту процедурных рисков и рекомендует сохранять основания решения и обеспечить реальную возможность человеческой проверки. Статистические выводы о дискриминации не делаются по одному заявителю. Для групповых выводов требуется проверяемая выборка, критерии сопоставимости и отдельная оценка обработки чувствительных данных.

Поле/контроль	Правило заполнения или рекомендуемое значение
Решение	Документированный отказ
AI-роль	Рекомендация или автоматическое решение
Процедура	Доступность объяснения
Пересмотр	Может ли сотрудник изменить результат
Вывод	Контекстная оценка, не обвинение без анализа

Проверка на практике

Для отказа в социальной услуге сначала устанавливается, была ли автоматизированная рекомендация обязательной для сотрудника. Если сотрудник мог её изменить, но не пользовался возможностью, это отличающийся риск организации процесса. Далее проверяется предоставление уведомления и доступность обжалования. На основании единственного обращения не рассчитываются групповые показатели дискриминации. Потенциальный вред можно убедительно описать и без ложного статистического обобщения.

Приложение 33 Сценарий: обновление модели без повторной оценки

Назначение

Условный пример: поставщик заменяет алгоритм, но организация продолжает ссылаться на предыдущий тестовый протокол. Старая оценка не обязательно становится бесполезной, однако её применимость зависит от характера изменений. Observatory требует сравнения версий, перечня затронутых функций, возможности повторения тестов и наличия новых защитных мер.

Порядок применения и ограничения

Если поставщик не раскрывает изменения, публикация должна указывать ограничение знания, а не утверждать, что модель стала опасной. Практическая рекомендация состоит в установлении заранее определённых триггеров повторной независимой оценки при изменении ключевых функций и контекста применения.

Поле/контроль	Правило заполнения или рекомендуемое значение
Версия A	Что ранее проверялось
Версия B	Какие функции изменились
Эквивалентность	Чем она подтверждена
Риск	Предел применимости старого теста
Действие	Повторный targeted review

Проверка на практике

Повторное тестирование после обновления модели определяется не календарём, а характером изменения: новый источник данных, иной алгоритм ранжирования, расширение сферы применения или изменение автономных функций. Для малозначительного изменения интерфейса может быть достаточно регрессионного теста, для новой модели — полного пересмотра контрольного сценария. Разработчик обязан подтвердить сведения в рамках применимого режима; независимый Observatory сообщает, какое подтверждение было доступно и какой части системы оно касается.

Приложение 34 Сценарий: журналистский всплеск AI-инцидентов

Назначение

Условный пример: после резонансного сообщения редакции начинают активнее писать об использовании алгоритмов, и количество найденных случаев за месяц удваивается. Без изменения знаменателя и охвата СМИ нельзя заключить, что реальная частота инцидентов выросла вдвое. Observatory раздельно показывает динамику уникальных событий, количество публикаций на событие и долю подтверждений официальными документами.

Порядок применения и ограничения

Рекомендуется проводить анализ чувствительности к источникам и языкам, а также маркировать период повышенного внимания СМИ. Вывод об общественном риске делается осторожно: обнаружено больше сообщений, но распространённость событий может оставаться неопределённой.

Поле/контроль	Правило заполнения или рекомендуемое значение
Наблюдение	Рост числа найденных статей
Дедупликация	Несколько статей об одном эпизоде
Охват	Изменилось внимание СМИ
Вывод	Нельзя вывести реальную распространённость
Мера	Показать альтернативные ряды

Проверка на практике

После медиавсплеска количество статей может вырасти, даже если число событий осталось постоянным. Поэтому анализ включает число уникальных эпизодов и отдельный показатель публикаций на эпизод. Особое внимание уделяется повторным сообщениям об одном судебном решении. Аналитик проверяет изменение языков и СМИ, попавших в мониторинг. В публичном резюме используется ограниченное утверждение: «зарегистрирован рост сообщений», а не «реальные инциденты удвоились».

Приложение 35 Паспорт ежегодного отчёта обсерватории

Назначение

Ежегодный публичный отчёт должен отражать не только достижения, но и ограничения: сколько сообщений получено, сколько проверено, сколько отвергнуто по недостатку данных, сколько исправлено после жалоб, какие конфликты интересов возникли и как они были разрешены. Показатели разделяются по секторам и географии лишь при достаточном объёме и правомерности публикации.

Порядок применения и ограничения

В документ включается независимая оценка методик, статус выполнения предыдущих рекомендаций и план следующего года. Нельзя приписывать собственной организации успех решений государственных органов, если её роль ограничилась направлением комментария. Все значимые утверждения об институциональных отношениях сопровождаются подтверждаемыми документами.

Поле/контроль	Правило заполнения или рекомендуемое значение
Методы	Версии и изменения за год
События	Разделение signals / hazards / incidents
Качество	Внешние проверки и исправления
Финансирование	Структура поддержки и риски
Влияние	Документированные реакции
План	Приоритеты и ограничения ресурсов

Проверка на практике

Ежегодный отчёт показывает долю карточек, где первоначальная гипотеза не подтвердилась. Это важный показатель честной проверки, а не повод скрыть неудачи. Сюда же включаются существенные замечания внешних рецензентов, доля исправленных ссылок и общие обстоятельства финансовых конфликтов. Реакцию регулятора или суда нельзя называть институциональной победой Observatory без доказанной роли исследования. Такой формат создаёт более устойчивую репутацию, чем подборка только положительных отзывов.



Приложение 36 Сводная матрица принятия решения о запуске

Назначение

Итоговая матрица применяется перед внешним объявлением запуска программы. Наличие красивого сайта и названия недостаточно, если отсутствуют правовая основа обработки данных, независимая проверка и механизм исправления. Для каждого элемента указывается документальное свидетельство, ответственный и решение о допустимости пилотного использования.

Порядок применения и ограничения

Статус «готово» присваивается только после фактического теста процесса. Если обеспечение конфиденциального архива не проверено, публичный канал индивидуальных обращений не открывается, даже если остальные элементы работают. Организация вправе начать с открытых агрегированных источников, пока чувствительные процедуры не готовы.

Поле/контроль	Правило заполнения или рекомендуемое значение
Право	Проверено юридическим специалистом
Независимость	Устав, отводы, раскрытие финансирования
Безопасность	Разделение хранения и доступов протестированы
Метод	Словарь и peer review утверждены
Исправления	Рабочий канал оспаривания
Пилот	Ограничения публично обозначены

Проверка на практике

Перед масштабированием требуется не демонстрация работающего интерфейса, а тест критических процедур с имитацией ошибок. Команда должна суметь удалить личные данные из публичного экспорта, восстановить архив, отстранить конфликтного эксперта и исправить опасную публикацию. Если тест провален, соответствующая функция пилота не открывается. Эта матрица допускает частичный запуск: например, мониторинг официальных открытых документов без приёма чувствительных обращений, пока конфиденциальный контур не проверен.

Приложение 37 Схема данных и связей между исследовательскими реестрами

Назначение

Предлагается модель данных, в которой объект AI-системы, контекст применения, свидетельство, событие, риск, рекомендация и публикация являются самостоятельными сущностями. Это предотвращает смешение коммерческого названия продукта с фактическим решением конкретного государственного или частного оператора. Одно событие может относиться к нескольким риск-сценариям, одно доказательство поддерживать несколько утверждений, а одна публикация включать много событий. Для каждого отношения указываются дата, автор подтверждения и его основание. Внутренние идентификаторы персональных заявителей не помещаются в открытые карточки.

Порядок применения и ограничения

Для реализации достаточна обычная реляционная система с журналом изменений, без обязательного применения машинного обучения. Минимальное правило: удалить ошибочную публичную строку можно только сохранив закрытую историю исправления. Смена версии AI-системы не переписывает старые инциденты. При миграции базы проверяется сохранность ссылок и чисел, а сам переход документируется. Открытая выгрузка включает идентификаторы, словарь полей, версии и разрешённые атрибуты, но не ключи связи с персональными контактами. Качество схемы оценивается тестом восстановления истории одной карточки по исходному документу.

Поле/контроль	Правило заполнения или рекомендуемое значение
SYSTEM	Паспорт изделия и версии
DEPLOYMENT	Использование с оператором и географией
EVENT	Событие и доказательственный статус
EVIDENCE	Исходный материал и контроль происхождения
RISK	Сценарий, тяжесть и меры
RECOMMENDATION	Адресат и статус проверки
PUBLICATION	Версия отчёта и ссылки
PERSONAL	Закрытый отдельный контур

Приложение 38 Согласованность классификации независимыми оценщиками

Назначение

Независимость классификатора нужно проверять не только на уровне инструкции, но и по результату работы разных аналитиков. Для пилота отбирается заранее установленная доля карточек, которую два специалиста кодируют отдельно: AI incident, AI hazard, предварительный сигнал либо не относящееся к AI событие. До сравнения специалисты не видят выбор друг друга. По завершении фиксируются совпадения, причины расхождения и

решение арбитра. Простая доля совпадений полезна как первоначальный показатель, но может завышать качество, когда одна категория значительно преобладает.

Рабочий учебный пример для двух бинарных классов: из 40 карточек оба оценщика отметили 16 как «инцидент», оба 16 как «не инцидент», а в восьми разошлись (четыре в одну и четыре в обратную сторону). Доля совпадения 32/40 = 80%; поскольку у каждого по 20 положительных и 20 отрицательных оценок, случайное ожидаемое согласие составляет 50%; κ = 0,60. Это условная матрица, а не фактический результат исследования. Для четырёх исходных категорий из данного приложения понадобится соответствующая многоклассовая процедура, а не механическое применение бинарного примера.

Порядок применения и ограничения

Для достаточной выборки возможно дополнить проверку коэффициентом согласованности с учётом случайных совпадений, указав выбранную статистическую модель и её ограничения. Коэффициент не применяется автоматически к слишком малому или специфическому набору карточек. Если два оценщика постоянно расходятся в понимании «причинной роли ИИ», корректируется не оценка конкретного лица, а сама инструкция с примерами граничных ситуаций. После изменения инструкции проводится повторное кодирование небольшой контрольной выборки. Все результаты рассматриваются как показатели качества процесса, а не как доказательство истинности каждой карточки.

Поле/контроль	Правило заполнения или рекомендуемое значение
Выборка	Предварительно определённый набор
Кодирование	Два независимых эксперта
Слепой режим	Без предварительного согласования оценок
Разногласие	Причина и решение арбитра
Метрика	Совпадение и, при условиях, скорректированный коэффициент
Обновление	Версия инструкции и повторная проверка
Отчёт	Доля неразрешённых случаев

Приложение 39 Проверка временных рядов и структурных разрывов

Назначение

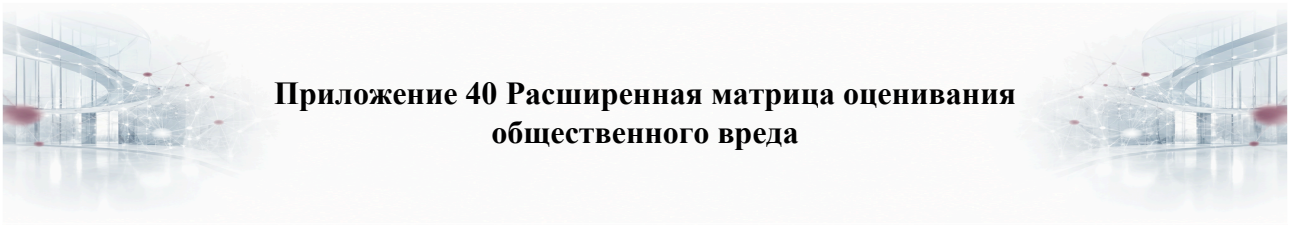
Увеличение количества зарегистрированных инцидентов может быть следствием расширения языков, добавления новых СМИ, введения формы обращения или изменения определения события. Поэтому временной ряд должен сопровождаться журналом изменений системы наблюдения. Если до июня использовались пять источников, а с июля двадцать, сравнение необработанных ежемесячных чисел не измеряет изменение распространённости событий. Метод может включать пересчёт исторического периода по новой схеме источников, но только если исходные материалы доступны для такой проверки.

Порядок применения и ограничения

Для каждого тренда публикуются абсолютные числа, доля подтверждённых первичными документами карточек, доля устранённых дублей и описание изменений охвата. При явном структурном разрыве данные не объединяются в одну непрерывную линию без оговорки. Скользящее среднее допустимо для визуализации, но не устраняет проблему отбора.

Заявление о значимом росте должно учитывать не только статистические интервалы, но и теоретический смысл измеряемой величины. При отсутствии правомерного знаменателя тенденция описывается как динамика зарегистрированных наблюдений, не реального вреда во всём обществе.

Поле/контроль	Правило заполнения или рекомендуемое значение
Временной период	Фиксированные границы
Каналы	Изменение языков и источников
Дедупликация	Единое правило события
Пересчёт	Возможен/невозможен и почему
Структурный разрыв	Пометка на каждом графике
Интерпретация	Наблюдаемая серия или реальная частота
Проверка	Чувствительность к смене охвата



Приложение 40 Расширенная матрица оценивания общественного вреда

Назначение

Общественный вред от ИИ следует описывать многомерно. Помимо прямых финансовых потерь учитываются последствия для здоровья, свободы передвижения, приватности, доступа к услугам, репутации, условиям труда и возможности эффективного обжалования. Наличие нескольких типов вреда не даёт права автоматически суммировать их как одинаковые единицы. Для сравнения применяется качественный профиль с отдельной оценкой тяжести, продолжительности, масштаба и обратимости. Публичная оценка относится к документированному воздействию системы, а не к моральной характеристике пострадавшего.

Порядок применения и ограничения

Для событий с недостаточной информацией предлагается явно выделять потенциальный вред и уже установленное последствие. Например, ошибочное санкционное совпадение может создать риск невозможности оплатить жильё, но если такого последствия не установлено, оно не записывается в графу фактического ущерба. При индивидуальных делах особенно важна проверка того, существуют ли эффективные средства смягчения: человеческий пересмотр, временный доступ к услуге, порядок исправления персональных данных. Результат анализа служит для приоритизации исследования, но не для расчёта компенсации без отдельной правовой и доказательной основы.

Поле/контроль	Правило заполнения или рекомендуемое значение
Тип вреда	Материальный, физический, процессуальный, социальный
Статус	Актуальный / потенциальный / не установлен
Тяжесть	Низкая / средняя / высокая / необратимая
Охват	Индивид / группа / инфраструктура
Продолжительность	Период последствий
Обратимость	Наличие действенной меры
Основание	Документ и пределы установленного

Приложение 41 Запрос поставщику ИИ о верифицируемости системы

Назначение

Шаблон запроса поставщику ориентирован на факты, необходимые для независимой проверки без требования безусловно раскрыть коммерчески охраняемый исходный код. Вопросы охватывают название и версию, назначение и ограничения системы, проведённые проверки, порядок обновлений, журналы событий, ответственность за данные и меры человеческого контроля. Поставщик может обоснованно ограничить доступ к определённым материалам, но отсутствие информации обязательно отражается как ограничение достоверности внешнего вывода. Ответ поставщика не превращается в независимый результат от одного факта получения.

Порядок применения и ограничения

При закрытой технической проверке стороны заранее определяют круг лиц, меры конфиденциальности и минимальный набор материалов, на основании которого можно подтвердить конкретный тезис. Нельзя подписывать соглашение, позволяющее поставщику по своему усмотрению переписывать или блокировать законный вывод независимого исследователя. Если договор не позволяет раскрыть сведения, достаточные для публичной проверки, публикуется строго ограниченное описание области исследования и остающейся неопределённости. После существенного обновления модели запрос повторяется в относящейся к изменению части.

Поле/контроль	Правило заполнения или рекомендуемое значение
Модель	Название, версия, дата и статус
Назначение	Предусмотренные и исключённые случаи
Данные	Категории и ответственность за качество
Тесты	Методы, выборка, ограничения
Обновления	Журнал существенных изменений
Логи	Доступный период и механизм сохранения
Контроль	Кто может остановить решение
Публикация	Условия независимого вывода

Приложение 42 Анализ чувствительности показателя с условным примером

Назначение

Предположим, в пилотном наборе содержится 100 сообщений: 55 подтверждённых, 25 предварительных и 20 исключённых как не относящихся к ИИ. Если исследователь объявит «80 инцидентов», он объединит достоверные случаи с неустановленными сигналами и завысит определённую им категорию. Корректно сообщить 55 подтверждённых событий и 25 ожидающих проверки сообщений; при этом оба числа характеризуют именно

исследовательский набор, а не всю экономику. Этот пример условный и не описывает результаты реального мониторинга ARGA.

Порядок применения и ограничения

Чувствительность проверяется альтернативными правилами. Если все предварительные сообщения позднее подтвердятся, число подтверждённых вырастет до 80; если ни одно не подтвердится, останется 55. Интервал 55–80 не является статистическим доверительным интервалом распространённости вреда: это диапазон исходов классификации в данном фиксированном наборе. Такой способ особенно полезен руководству для оценки надёжности выводов. При публикации необходимо также проверить устранение дублей и правомерность исходных данных.

Поле/контроль	Правило заполнения или рекомендуемое значение
Подтверждено	55 условных карточек
Предварительно	25 условных карточек
Исключено	20 условных карточек
Минимум при текущей классификации	55 подтверждённых
Максимум при подтверждении сигналов	80 подтверждённых
Нельзя утверждать	Доля инцидентов в общей популяции
Оговорка	Сценарий иллюстративный, не эмпирический результат

Приложение 43 Порядок обработки запроса об удалении или ограничении данных

Назначение

Поступивший запрос субъекта данных требует отделения права на исправление персональных сведений от спора с содержанием научного вывода. В ЕС применимость прав GDPR зависит от роли организации, основания и цели обработки, категории данных и предусмотренных исключений. Нельзя обещать универсальное удаление всех следов исследования или автоматически отказать под предлогом публичного интереса. Ответственный специалист регистрирует запрос, проводит юридическую проверку и определяет, где хранятся оригиналы, рабочие копии и публичные версии.

При ответе на запрос фиксируются личность и полномочия заявителя, роль организации, вид данных, применимое право и индивидуально обоснованный результат. Право на стирание не является абсолютным, но исключения также нельзя применять автоматически по ссылке на «исследования». Особо чувствительные материалы и сведения об уголовных правонарушениях проходят отдельную проверку. При исправлении опубликованной ошибки версия исследования и уведомление внешних получателей рассматриваются отдельно от правомерности сохранения внутренних доказательств. [13]

Порядок применения и ограничения

При удовлетворении требования исправление проводится на всех контролируемых поверхностях: в базе, опубликованном файле, индексах и доступных копиях, насколько это технически и юридически возможно. Если существует законное основание сохранить

некоторые сведения, это объясняется заявителю в предусмотренном порядке. Параллельно оценивается, не станет ли само описание удаления новым раскрытием чувствительного дела. Персональные данные в журналах безопасности обрабатываются отдельно в соответствии с установленными сроками и основаниями. Результат не должен зависеть от того, является ли лицо сторонником Observatory или его критиком.

Поле/контроль	Правило заполнения или рекомендуемое значение
Тип запроса	Доступ / исправление / удаление / ограничение
Контролёр	Ответственный субъект обработки
Основание	Норма и возможные исключения
Объекты	База / архив / публикация / копии
Решение	Исполнено / частично / мотивированный отказ
Срок	Применимая юридическая процедура
Контроль	Подтверждение исполнения и журнал версии



Приложение 44 Модель распределения ответственности при AI-инциденте

Назначение

Схема предназначена для поиска фактов, а не для автоматического распределения юридической вины. Она различает разработчика общей модели, поставщика прикладной системы, оператора развертывания, владельца данных, сотрудника, принимающего юридически значимое решение, и возможные надзорные органы. В конкретном деле роли могут совпадать, а договорная цепь может включать международных подрядчиков. Каждое утверждение об ответственности нуждается в применимом праве, документированном действии и причинной связи.

Порядок применения и ограничения

При анализе ошибочного банковского решения важно установить, была ли модель лишь источником сигнала, какие сведения ввёл банк и мог ли его сотрудник изменить результат. При государственной автоматизации аналогично проверяются компетенция органа и процессуальные гарантии. Observatory вправе построить карту потенциальной ответственности и рекомендовать дополнительную проверку. Однако определение состава правонарушения, меры наказания или гражданско-правовой компенсации остаётся предметом компетентной юридической процедуры.

Поле/контроль	Правило заполнения или рекомендуемое значение
Разработчик	Доказанная роль базовой технологии
Поставщик	Интеграция, документация, обновления
Деплойер	Контекст и цели применения
Оператор	Фактическое юридическое решение
Данные	Происхождение и корректность
Право	Юрисдикция и применимые обязанности
Вывод	Гипотеза об ответственности, не приговор

Приложение 45 План независимой методической консультации

Назначение

Цель консультации с исследователями, практикующими юристами и представителями затронутых групп — испытать саму методику Observatory до широкого применения. Предварительно публикуются рабочие определения инцидента, риска, источника и показателя; участники получают конкретные вопросы и возможность обосновать замечания. Состав консультации раскрывается по профессиональным ролям при соблюдении прав на персональные данные. Коммерческие разработчики могут участвовать на прозрачных условиях, но не получают право утверждать итоговые методические правила.

Порядок применения и ограничения

После окончания обсуждения секретариат готовит таблицу комментариев и мотивированных ответов: что принято, что отклонено и какие аспекты требуют исследования. Методические изменения сопровождаются номером версии, а при их существенном влиянии пересчитываются тестовые результаты пилота. Консультация не выдаётся за официальное участие международной организации, если оно не подтверждено её собственными документами. Принцип независимости означает открытость аргументам, а не обязанность уравнивать достоверность каждого мнения.

Поле/контроль	Правило заполнения или рекомендуемое значение
Предмет	Конкретный раздел методики
Участники	Разные дисциплины и группы
Конфликты	Раскрытие релевантных отношений
Вопросы	Определения, исключения и риски
Ответ	Принят / частично / мотивированно отклонён
Версия	Дата изменения протокола
Публичность	Отчёт без необоснованных институциональных заявлений

Приложение 46 Программа независимой оценки безопасности архива

Назначение

Перед приёмом индивидуальных обращений проверяется информационная безопасность: создание учётных записей, выдача и отзыв доступа, резервирование, восстановление, экспорт, аудиторский журнал и безопасное удаление. Тестирование проводится в контролируемой среде без использования реальных дел в качестве тренировочных файлов. Публичная форма не должна напрямую давать доступ к исходному закрытому архиву, даже если обе системы обслуживаются одним провайдером. Доступ переводчиков и внешних reviewers ограничивается только назначенными материалами.

Порядок применения и ограничения

Отчёт об испытании содержит сценарий, ожидаемый безопасный результат, фактический итог и корректирующую меру. Критическими считаются возможность неконтролируемого чтения чужого досье, восстановление маскированного имени из опубликованного файла и потеря оригинальных доказательств после аварии. Такие дефекты требуют исправления до запуска соответствующей функции, а не обещания исправить «по ходу работы». Последующие тесты повторяются после обновления системы и смены поставщика хранения.

Поле/контроль	Правило заполнения или рекомендуемое значение
Доступ	Роли и отзыв учётной записи
Загрузка	Изоляция и сканирование вложений
Маскирование	Невосстановимость скрытых данных
Резерв	Фактический тест восстановления
Экспорт	Отсутствие чувствительных метаданных
Журнал	Кто видел и изменял материалы
Критерий запуска	Нет открытых критических дефектов

Приложение 47 Метод проверки институциональных заявлений о сотрудничестве

Назначение

Публичная аналитическая организация должна быть особенно аккуратна с названиями ООН, UNESCO, OECD, Совета Европы и национальных регуляторов. Для каждого заявления о сотрудничестве требуется установить конкретное действие и документ: регистрация на мероприятии, направленный комментарий, официальное приглашение, совместная работа либо аккредитационный статус. Разные формы не заменяют друг друга. Указание названия международного института в библиографии не создаёт партнёрства с ним.

Порядок применения и ограничения

Редактор готовит таблицу «заявленная формулировка — доказательство — допустимое выражение». Например, наличие уведомления о получении письменного комментария позволяет написать «материал направлен и получен», но не «позиция одобрена». Совместная публикация допускает указание фактических соавторов, но логотип партнёра используется только на согласованных условиях. После истечения срока аккредитации либо завершения проекта веб-страница обновляется. Такая процедура снижает риск репутационной ошибки и защищает независимость исследования от институционального маркетинга.

Поле/контроль	Правило заполнения или рекомендуемое значение
Факт	Событие с датой и предметом
Доказательство	Письмо, реестр, программа, контракт
Уровень	Подача / участие / соавторство / статус
Логотип	Только с подтверждённым правом использования
Формулировка	Не преувеличивать отношения
Актуальность	Перепроверка срока
Контроль	Независимый редактор

Приложение 48 Инструмент контроля межъязыковой правовой терминологии

Назначение

При международном исследовании важные ошибки возникают не из-за отсутствия источника, а из-за неправильного перевода процессуального термина. Слова warrant, notice, alert, suspicion, investigation, sanction, delisting и их аналоги имеют неодинаковую юридическую нагрузку в разных системах. Поэтому переводческий протокол связывает исходный термин с контекстом правопорядка, актом и переводом на язык доклада. Нельзя заменять неоднозначное иностранное понятие русским термином, создающим более сильное юридическое утверждение.

Порядок применения и ограничения

Для критических абзацев рекомендуется обратная проверка специалистом языка и юристом соответствующей юрисдикции. Машинный перевод может использоваться в качестве черновика, но не как независимый юридический источник. Если понятие не имеет прямого аналога, сохраняется термин оригинала с пояснением. Журнал исправлений фиксирует случаи, когда перевод менял правовой смысл опубликованного вывода. Системный контроль терминологии особенно важен для судебных актов, документов о санкциях и форм международного полицейского обмена.

Поле/контроль	Правило заполнения или рекомендуемое значение
Исходный термин	Текст и юрисдикция
Документ	Вид акта и дата
Предлагаемый перевод	Русское обозначение
Предел	Чего термин НЕ означает
Проверяющий	Юрист/переводчик
Версия	Дата словаря
Исправление	Риск изменения смысла

Приложение 49 Метод аудита собственных ошибок AI Observatory

Назначение

Observatory должен проводить независимую оценку не только внешних систем, но и собственных публикаций. Раз в год отбирается выборка карточек, включая подтверждённые случаи, отклонённые гипотезы и материалы, которые вызвали жалобы. Вторая команда восстанавливает первичные источники, классификацию, расчёты, выполнение правила права на ответ и проверку приватности. Признаки ошибок делятся на технические, фактические, статистические, правовые и этические. Каждая категория требует собственного способа исправления, поэтому единое число «ошибок» без тяжести мало информативно.

Порядок применения и ограничения

Результат аудита публикуется с указанием охвата проверки и ограничений. Проблемы, связанные с прямым ущербом человеку, получают срочный статус и отдельный процесс исправления. Редакция должна иметь возможность не соглашаться с рецензентом, но только с мотивированным ответом, включённым в закрытый журнал. Сокрытие ошибок ради репутации прямо противоречит мандату общественного контроля. При этом добросовестное изменение гипотезы на основании новых данных не приравнивается к нарушению исследовательской этики.

Поле/контроль	Правило заполнения или рекомендуемое значение
Выборка	Заранее установленная и включающая спорные случаи
Проверка	Источники, выводы, право, безопасность
Категория	Техническая / фактическая / статистическая / правовая
Тяжесть	Есть ли риск вреда
Реакция	Исправление, переоценка, отзыв
Публичность	Агрегированный отчёт с методикой
Обучение	Изменение протокола и контроль повторения

Приложение 50 План устойчивости при прекращении финансирования

Назначение

Независимость зависит от способности организации продолжить хранение доказательств, исправление опубликованных ошибок и защиту обращающихся даже при прекращении крупного гранта. Для этого в план включаются резервные процедуры хранения, права на экспорт данных, правила завершения договоров и минимальный круг ответственных лиц. Нельзя принимать чувствительное обращение, если при закрытии проекта материалы могут остаться доступными коммерческому поставщику без ясного правового основания.

Порядок применения и ограничения

При стресс-сценарии сокращения доходов сначала прекращается расширение мониторинга, затем пересматривается тематика новых исследований, но сохраняются критические функции по защите данных и исправлению старых публикаций. Права на интеллектуальные результаты и архив определяются до получения гранта. Решение о закрытии программы должно включать уведомление заявителей, правомерную передачу либо удаление сведений и публичное объяснение судьбы методик. Эта мера делает институциональную независимость не декларацией, а проверяемой устойчивостью.

Поле/контроль	Правило заполнения или рекомендуемое значение
Событие	Потеря значительного финансирования
Резерв	Минимальные расходы на безопасность и исправления
Данные	Право на безопасный экспорт или удаление
Люди	Ответственный за обращения после остановки
Метод	Доступность опубликованного словаря
Закрытие	Соблюдение хранения и уведомлений
Контроль	План проверяется до запуска

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.

20.01 Задавать исследовательский вопрос до сбора
Раздел 1. Результат: Вывод отвечает заданному вопросу и не выходит за наблюдаемую совокупность.

20.02 Разделять наблюдение и властное решение
Раздел 2. Результат: Публичный текст точно передаёт статус исследования.

20.03 Переносить институциональный опыт с учётом ресурсов
Раздел 3. Результат: Пилот выполним имеющимся составом без неподтверждённых заявлений о масштабе.

20.04 Проверять правовое основание мониторинга
Раздел 4. Результат: Публичная карточка содержит только обоснованно раскрываемые данные.

20.05 Разделять систему и её применение
Раздел 5. Результат: Событие связано с фактическим оператором и назначением.

20.06 Назначать разные роли проверки и публикации
Раздел 6. Результат: Существенная публикация имеет вторую документированную проверку.

20.07 Защищать право на неудобный вывод
Раздел 7. Результат: Решение о публикации принято по доказательствам и утверждённой процедуре.

20.08 Раскрывать зависимость финансирования
Раздел 8. Результат: Финансирование и меры управления зависимостью доступны для проверки.

20.09 Применять отвод по конкретному вопросу
Раздел 9. Результат: Заинтересованный участник не утверждает вывод по своему вопросу.

20.10 Сопоставлять страны при одинаковом охвате
Раздел 10. Результат: Межстрановой вывод сопровождается проверкой наблюдаемости.

20.11 Фиксировать каналы мониторинга
Раздел 11. Результат: График отмечает смену охвата и не выдаёт её за рост риска.

20.12 Разделять приём сообщения и подтверждение
Раздел 12. Результат: Карточка содержит статус проверки и следующий конкретный запрос.

20.13 Различать сигнал опасность и инцидент
Раздел 13. Результат: Категория подтверждена наблюдаемыми признаками, неопределённость сохранена.

20.14 Определять смысл показателя до расчёта
Раздел 14. Результат: Показатель имеет определение, единицу и воспроизводимый расчёт.

20.15 Указывать знаменатель и механизм отбора
Раздел 15. Результат: Внешний текст содержит корректный знаменатель и ограничения отбора.

20.16 Прослеживать источник каждого существенного факта
Раздел 16. Результат: Карточка показывает число независимых оснований, а не перепечаток.

20.17 Предоставлять содержательную возможность ответа
Раздел 17. Результат: Позиция адресата рассмотрена по существу и отражена в версии публикации.

20.18 Связывать риск с мерой и владельцем
Раздел 18. Результат: У риска есть ответственная мера и подтверждение её проверки.

20.19 Обосновывать приоритет по последствиям
Раздел 19. Результат: Очередность рассмотрения объяснена по единой шкале.

20.20 Давать рекомендацию с исполнимым действием
Раздел 20. Результат: Рекомендация содержит действие, адресата и критерий исполнения.

20.21 Проверять косвенную идентификацию
Раздел 21. Результат: Открытая версия не раскрывает лишних идентифицирующих сочетаний.

20.22 Учитывать доступность процедуры участия
Раздел 22. Результат: Доступность улучшена, сравнение периодов скорректировано.

20.23 Минимизировать международный обмен
Раздел 23. Результат: Передача документирована по составу, цели и получателю.

20.24 Проверять санкционное совпадение отдельно от риска
Раздел 24. Результат: Публикация не приписывает лицу неподтверждённый санкционный статус.

20.25 Разделять официальную запись и автоматическую сводку
Раздел 25. Результат: В карточке раздельно указаны источник, лицо и доказанный статус.

20.26 Не приравнивать AML-сигнал к преступлению
Раздел 26. Результат: Текст сохраняет различие между риском, решением и юридической квалификацией.

20.27 Проверять доступность пересмотра публичного решения
Раздел 27. Результат: Рекомендация адресована владельцу процедуры и проверяется учебным обращением.

20.28 Включать проверяемые условия в закупку
Раздел 28. Результат: Контрактный контроль связан с конкретными свидетельствами.

20.29 Не переносить общий рейтинг на развёрнутый продукт
Раздел 29. Результат: Вывод относится к фактически проверенному режиму применения.

20.30 Отделять происхождение контента от его влияния
Раздел 30. Результат: Публикация не подменяет доказательство влияния предположением об авторстве.

20.31 Сопоставлять общественные издержки по определённой базе
Раздел 31. Результат: Сравнение воспроизводимо либо ограничения явно препятствуют ранжированию.

20.32 Согласовывать ожидания заявителя
Раздел 32. Результат: Заявитель понимает статус участия и способы дальнейшего обращения.

20.33 Публиковать материал по степени подтверждённости
Раздел 33. Результат: Заголовок и содержание соответствуют доказательственному статусу.

20.34 Сохранять историю исправлений
Раздел 34. Результат: Пользователь видит, что изменено, почему и какая версия актуальна.

20.35 Проводить независимую рецензию ключевых выводов
Раздел 35. Результат: Существенные расхождения разрешены либо отражены как ограничения.

20.36 Проверять восстановление и разграничение доступа
Раздел 36. Результат: Тест восстановления и отказа неразрешённого доступа завершён протоколом.

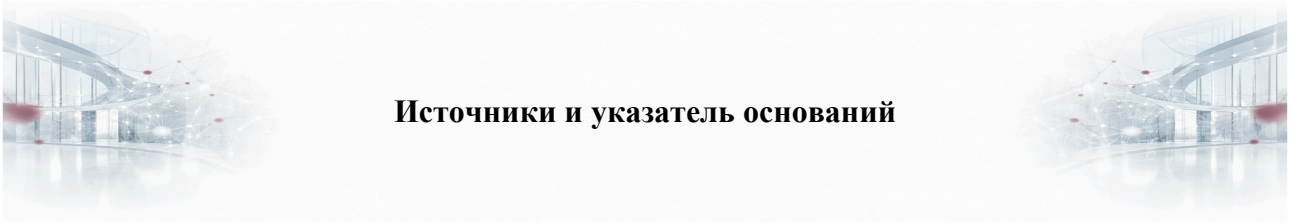
20.37 Измерять качество собственной работы
Раздел 37. Результат: Отчёт показывает качество и ограничения оценки наряду с объёмом.

20.38 Подтверждать институциональные заявления документами
Раздел 38. Результат: Каждое заявление о сотрудничестве имеет соответствующее основание.

20.39 Пилотировать ограниченный корпус до масштабирования
Раздел 39. Результат: Решение о расширении основано на качестве процедуры и выявленных пробелах.

20.40 Соотносить обещания с кадровым ресурсом
Раздел 40. Результат: План соответствует доступному времени с сохранением контроля качества.

20.41 Принимать решение о запуске по подтверждённой готовности
Раздел 41. Результат: Решение содержит подтверждения готовности, ограничения и условия пересмотра.



Источники и указатель оснований

Источники сгруппированы по их предметной применимости. Для каждого приведена прямая ссылка на официальный текст либо страницу издателя. Ключевые нормативные и методические документы сверены на дату подготовки; ссылки на внешние ресурсы могут меняться, а применимость конкретной нормы требует самостоятельной проверки перед использованием. Академические или добровольные методики не приравниваются к обязательным правовым требованиям.

[1] ООН. Independent International Scientific Panel on AI: FAQ и полномочия группы. https://www.un.org/independent-international-scientific-panel-ai/en/faq

[2] ООН. Global Dialogue on AI Governance: задачи и формат. https://www.un.org/global-dialogue-ai-governance/en

[3] UNESCO. Recommendation on the Ethics of Artificial Intelligence. https://www.unesco.org/ethics-ai/en/recommendation

[4] UNESCO. AI Readiness Assessment Methodology (RAM). https://www.unesco.org/ethics-ai/en/ram

[5] UNESCO. Ethical Impact Assessment (EIA). https://www.unesco.org/ethics-ai/en/eia

[6] OECD. AI Principles, редакция 2024 года. https://www.oecd.org/en/topics/ai-principles.html

[7] OECD.AI. AI Incidents and Hazards Monitor: методология и ограничения. https://oecd.ai/en/incidents-methodology

[8] OECD. Towards a common reporting framework for AI incidents (2025). https://www.oecd.org/en/publications/towards-a-common-reporting-framework-for-ai-incidents_f326d4ac-en.html

[9] Совет Европы. HUDERIA: методология оценки влияния ИИ на права и демократию. https://www.coe.int/en/web/artificial-intelligence/huderia-risk-and-impact-assessment-of-ai-systems

[10] Совет Европы. Рамочная конвенция об ИИ (CETS 225): статус договора. https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=225

[11] EU AI Act. Консолидированный Регламент (ЕС) 2024/1689, 27.07.2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02024R1689-20260727

[12] Регламент (ЕС) 2026/1744: Digital Omnibus on AI. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32026R1744

[13] GDPR. Регламент (ЕС) 2016/679. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[14] NIST. Artificial Intelligence Risk Management Framework 1.0. https://www.nist.gov/itl/ai-risk-management-framework

[15] NIST. AI RMF Generative AI Profile (AI 600-1). https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

[16] ISO/IEC 42001:2023. AI management systems. https://www.iso.org/standard/42001

[17] ISO/IEC 23894:2023. Guidance on AI risk management. https://www.iso.org/standard/77304.html

[18] OHCHR. UN Guiding Principles on Business and Human Rights. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

[19] International AI Safety Report 2026. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026

[20] OECD.AI. Policy Observatory portal and dashboards. https://oecd.ai/en/dashboards/ai-principles/overview

[21] UNESCO. Country profiles and Global Hub. https://www.unesco.org/ethics-ai/en/global-hub

[22] ООН. Независимая научная группа по ИИ: официальная страница. https://www.un.org/independent-international-scientific-panel-ai/en

[23] EUR-Lex. Rules for trustworthy AI in the EU: даты применения. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:4762484

Примечание о статусе исследования

Доклад подготовлен как независимое аналитическое и методологическое предложение Observatoire ARGA. Он не свидетельствует об официальном одобрении модели структурами ООН, UNESCO, OECD, Совета Европы, NIST или органами Европейского союза; никакой их статус не подразумевается без отдельного подтверждения. Практическое применение положений требует дальнейшей юридической, технической и институциональной проверки.

Научные и методологические источники

[M1] Krippendorff K. Reliability in Content Analysis: Some Common Misconceptions and Recommendations. Human Communication Research, 30(3), 2004, pp. 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

[M2] OECD/European Union/EC-JRC. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, 2008. https://doi.org/10.1787/9789264043466-en

[M3] Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 1960, pp. 37–46. https://doi.org/10.1177/001316446002000104