

АВТОРСКАЯ АНАЛИТИЧЕСКАЯ СТАТЬЯ

Редакционная дата: 31 октября 2026 г.

Запоздалая подпись: кому принадлежит решение, принятое машиной?

> Летом 2020 года экзаменационная оценка выглядела как математический ответ на педагогический вопрос. Но поиск того, кто «принял» это решение, привёл к цепочке людей: каждый обладал частью полномочий, и никто не видел всей картины. Это не только история об ошибке компьютера, но и история организации, которая разделила решение, а ответственность забыла собрать заново.

Оценка, которую ученик не писал в экзаменационной работе

13 августа 2020 года в Англии объявили результаты экзаменов продвинутого уровня **A-level**, от которых зависит поступление тысяч учеников в университет. Но тем летом экзаменов не было: 18 марта из-за пандемии Covid-19 британское правительство отменило испытания. Взамен ввели механизм, использовавший оценки школ, составленный ими порядок учеников и статистическую обработку, призванную согласовать распределение результатов с прежними годами.[1][2]

Полученная учеником отметка была определённой, без половины буквы. Путь к ней оказался распределён: учителя представили оценки и порядок учеников; регулятор разработал метод стандартизации; экзаменационные комиссии выпустили результаты

по формальным требованиям; министерство задало цель политики. Робот не занимал место экзаменатора. Не существовало самой письменной работы, которую можно было бы проверить. Государство пыталось заменить образовательное событие, которого не произошло.

Через четыре дня, 17 августа, правило изменили. В применимых случаях итогом стала более высокая из двух величин: школьная или расчётная оценка. Позднейший анализ Ofqual охватил определённую выборку 18-летних учеников Англии — **457 420 предметных записей** для **246 110 учеников**. Расчёт совпал со школьной оценкой примерно в **59% записей**, оказался ниже в **39%** и выше примерно в **2%**. Это доли предметных записей, а не учеников; утверждать, что оценки снизили 39% учеников, было бы неточно.3

Расхождение с оценкой учителя не доказывает, что учитель был прав или что компьютер обошёлся с учеником несправедливо в каждом случае. Оно доказывает другое: система вынесла суждение с серьёзным личным последствием, а через несколько дней организация изменила основание этого суждения. Кому принадлежало первое решение? Кто нёс его последствия? Кто мог остановить его до того, как оно дошло до ученика?

2 сентября 2020 года председатель Ofqual Роджер Тейлор (Roger Taylor) написал парламентскому комитету по образованию, что регулятор **полностью принимает свою долю ответственности**. Решение использовать статистически стандартизированные учительские оценки следовало министерскому указанию, а за исполнение отвечал регулятор. Вывод был примечателен: улучшения одного алгоритма было недостаточно, чтобы сделать результаты приемлемыми; частью проблемы были ограничения данных и природа самого процесса.4

Легко назвать произошедшее «алгоритмическим скандалом» и пойти дальше. Такое название странным образом оправдывает людей. Уравнение не выпускает министерское указание, не решает, что справедливость равна сходству распределений двух лет, и не выбирает допустимый ущерб, если индивидуальный результат расходится с ожиданием ученика. Расчёту предшествовали человеческие решения.

Машина может исполнить выбор. Она не стирает человеческую историю этого выбора.

Четыре вида власти, скрытые в одном слове

Организации говорят, что «решение принято», словно у действия есть один очевидный владелец. Но институциональное решение — не точка, а движение между разными правами: предложить действие, утвердить правило выбора, санкционировать результат, остановить или исправить его.

В статье «Разделение собственности и контроля» (Separation of Ownership and Control) 1983 года Юджин Фама (Eugene F. Fama) и Майкл Дженсен (Michael C. Jensen) различили **управление решением (Decision Management)** — предложение и исполнение вариантов — и **контроль решения (Decision Control)** — их утверждение и наблюдение за результатом. Они изучали устройство организаций и конфликт между собственниками и менеджерами, а не ИИ. Но различие помогает увидеть то, что скрывает современный результат, появляющийся одним нажатием.5

Возьмём **условный** банк. Модель оценивает вероятность дефолта заявителя. Техническая команда создаёт оценку; риск-служба выбирает порог принятия; руководство утверждает допустимый убыток; операционная система отправляет уведомление; поддержка объясняет результат; лицензированный банк остаётся стороной правовых и договорных отношений с клиентом. Кто отказал в кредите?

На два разных вопроса можно дать два правильных ответа. **Внешняя юридическая ответственность** определяется применимым правом, договором и фактами и не исчезает из-за покупки модели у поставщика. **Внутренняя причинная способность** может принадлежать тому, кто утвердил переменную, задал порог или разрешил работу без проверки. Одно описание не заменяет другое: серьёзному расследованию нужны оба.

Уровень	Вопрос	Необходимое полномочие	Что его не доказывает
Определение цели	Какой интерес система вправе оптимизировать?	Утвердить цель, ограничения и права, не подлежащие обмену	Покупка «высокоточной модели»
Построение средства	Какие данные, допущения и издержки превращают цель в расчёт?	Проверять модель, альтернативы и пределы применимости	Показ коэффициентов уверенности
Придание силы	Когда рекомендация становится решением с последствием?	Утверждать пороги, полномочия и пределы автоматизации	Автоматическая подпись подразделения
Исправление и возмещение	Кто может остановить, отменить или пересмотреть последствие?	Действенное полномочие, время, ресурсы и путь обжалования	Ящик для жалоб, владелец которого ничего не меняет

Это не новый перечень юридических обязанностей для какой-либо юрисдикции, а предлагаемая операционная карта их отслеживания. Сбой возникает, когда четыре власти разделены без ясных соединений: проверяющий модель не вправе изменить цель; утвердивший цель не понимает уравнение; встречающий клиента может лишь прочесть заранее сформулированную причину.

Письменное полномочие не всегда является реальной властью

В статье Филиппа Агьона (Philippe Aghion) и Жана Тироля (Jean Tirole) 1997 года «Формальная и реальная власть в организациях» (Formal and Real Authority in Organizations) сформулирована простая глубокая мысль: формальное право решить не гарантирует практической способности выбрать, особенно когда подчинённый располагает сведениями, которые руководитель не успевает усвоить.6

Сложные модели расширяют этот разрыв. Директор по рискам может утверждать политику, не зная, как удаление переменной отразится на тысячах случаев. Специалист по данным знает эффект, но не вправе решать, приемлем ли он в кредитном или правовом смысле. Фраза «мы утвердили модель» может обозначать просмотр презентации, понимание риска или принятие последствий. Первое не доказывает второго и третьего.

Затем появляется проблема принципала и агента (Principal–Agent Problem). Совет делегирует исполнение руководству, руководство — техническую работу команде, команда может отдать поставщику размещение или эксплуатацию. Каждое звено преследует законную местную цель: сдать вовремя, уменьшить дефолты, сократить ответ, повысить одобрение. Согласие по местным целям не гарантирует решения в интересах, ради которых существует организация.

Формальный владелец полномочия может последним узнать, что это полномочие позволяет. Не обязательно из-за сговора или небрежности: так действует неравномерное распределение знания, времени и средств влияния. Если организация намеренно не возвращает реальную власть тому, кто несёт ответственность, «человеческая проверка» превращается в подпись, пришедшую после того, как отмена стала дорогой.

Когда уравнение правильно, а вопрос неверен

Стабилизация распределения оценок 2020 года относительно прежних лет не была бессмысленной целью. Без корректировки школы могли проявить разную щедрость, оценки — раздуться, сравнение поколений — исказиться. Ofqual прямо указывал на эти риски. Но статистическая стандартизация превратила различия между школами в последствие для отдельного ученика, который не мог показать знания на настоящем экзамене.14

Модель могла соответствовать заданной цели, тогда как сама цель была недостаточным основанием приемлемого решения. Между **валидностью расчёта внутри модели** и **правомерностью его применения к человеку** лежит не скорость процессора, а различие в первоначальном вопросе.

Случай показывает риск, плохо различимый из технической команды. Разработчики могут измерять стройность распределения; правительство — непрерывность образования; университеты — готовность к поступлению; ученики — справедливую возможность показать знания. Это не разные меры одного явления, а конфликтующие интересы и права. Одна целевая функция не примирит их автоматически. Порядок приоритетов и допустимая жертва требуют явного политического или институционального решения.

После кризиса британское Управление регулирования статистики опубликовало 2 марта 2021 года обзор. Оно потребовало различать ответственность заказчика модели, задающего цель, и разработчика; оценивать пригодность и общественную приемлемость; заранее определять основания и сроки оспаривания. Рекомендация состояла не в отказе от моделей, а в отказе представлять общественное решение как чисто техническое упражнение.7

Что в действительности решает модель в кредитной заявке?

Только для иллюстрации представим банк, использующий модель вероятности погашения $p(x)$, где x — доступные сведения заявления. Пусть B — дополнительный доход при погашении, L — убыток при дефолте, C — стоимость обслуживания. Упрощённое нейтральное к риску правило принимает заявку при следующем условии:

$$p(x)B - (1 - p(x))L - C > 0.$$

Это не настоящее банковское правило и не кредитный совет. Реальное кредитование сложнее и ограничено ликвидностью, капиталом, регулированием, равным обращением, модельным риском и правами клиента. Уравнение показывает одно: модель может дать **прогноз**, но не обязательно определяет B, L, C или право учреждения задавать цель именно так. Выбор одной «ожидаемой прибыли» сам по себе спорен.

Фраза «прогнозирует погашение» скрывает вопрос знания. Прогноз — не обещание; вероятность погашения не равна нравственной способности исполнить обязательство и не обязательно означает правовую допустимость финансирования. Нерепрезентативные данные или кредитные истории, сформированные старыми

стандартами, могут переносить прошлые решения в новое число. Признание пределов не устраняет необходимость решать; оно меняет допустимые утверждения об основании.

Представим оценку 0,79 при управленческом пороге 0,80. Аналитик не вправе сказать клиенту, что 0,79 было «приказом модели». Последствие создало человеческое решение. При изменении стоимости капитала или риск-бюджета порог сдвинется, хотя модель и клиентские данные останутся прежними. Организация создаёт новые решения политическим регулятором и ошибочно приписывает их системному «интеллекту».

Это важно заявителю, которому нужна проверяемая и исправимая причина; риск-службе, объясняющей чужую конструкцию; акционерам, которым нужно знать, кто изменил склонность к риску; надзору, для которого имени модели в реестре недостаточно.

Почему «мы не понимаем чёрный ящик» — не оправдание

Американские кредитные нормы дают полезное различие **только в пределах своей юрисдикции**. Раздел **§1002.9 Regulation B** в действующей опубликованной редакции регулирует уведомления о неблагоприятном решении и конкретные причины при условиях, зависящих от вида заявления. Официальное толкование говорит, что одного сообщения о недостижении нужного балла недостаточно: причины должны относиться к реально использованным факторам.8

В 2022 году Бюро финансовой защиты потребителей заявило в циркуляре, что непрозрачность алгоритма не оправдывает отсутствие причин. Этот циркуляр и связанный документ 2023 года были **отозваны 12 мая 2025 года**. Нельзя выдавать его прежнее толкование за действующую политику. Здесь важно отличить отозванное руководство от продолжающей действовать нормы, а не судить банк за пределами США.9

Методы атрибуции признаков могут объяснять прогноз модели, но объяснение расчёта не всегда объясняет институциональное решение. Первое указывает вклад исходных данных по определённому методу; второе должно объяснить выбор числа как стандарта, представление клиентских данных и политику, придавшую результату силу. Приближённые методы могут расходиться и не всегда описывают причинные связи.

И наоборот, система может раскрыть все вычислительные причины, а решение останется неприемлемым. Кто разрешил использовать исторические данные с этой целью? Кто решил, что цена ложного отказа полностью ложится на клиента?

Прозрачность — важное окно, но незаконное решение не становится законным. Знать, почему дверь закрылась, не значит иметь ключ.

Институциональная ошибка «человека, который утверждает»

Многие системы обещают человека в контуре (human in the loop). Это может быть полезно, но название ничего не сообщает о власти человека. Видит ли он факты, понимает ли пределы модели, может ли без скрытого наказания отступить от рекомендации, располагает ли временем и останавливается ли последствие до конца проверки?

Надзор не создаётся галочкой «утвердить». Перед сотнями однотипных рекомендаций сотрудник превращается в печать. Организация вставляет человека, чтобы доказать его присутствие, а работу строит так, что несогласие невозможно.

Добровольная американская **Рамочная система управления рисками ИИ (AI Risk Management Framework)** NIST 2023 года обращается к настоящей ответственности: **GOVERN 2.1** требует документировать роли и линии связи, **GOVERN 2.3** возлагает на высшее руководство ответственность за решения о рисках разработки и внедрения. Но рамка **добровольна** и не является законом, дарующим отсутствующие полномочия.10

Статья 14 **Акта ЕС об искусственном интеллекте (EU AI Act)** требует, чтобы надзор за высокорисковыми системами с учётом контекста понимал возможности и пределы, мог игнорировать или отменять результат, вмешиваться или остановить систему. Применение вводится поэтапно. По данным Еврокомиссии после поправок 2026 года, положения для ряда применений Приложения III начинают действовать **2 декабря 2027 года**, для систем в регулируемой продукции — **2 августа 2028 года**. Неверно считать это нынешней обязанностью саудовского банка или распространять каждую норму на любую модель.11

За пределами географии урок таков: человеческий надзор — не место сотрудника на схеме, а **практическая сила возразить до того, как возражение утратит ценность**.

Саудовская Аравия: от распределения обязанностей к их проверке

Ключевые принципы корпоративного управления финансовыми организациями, поднадзорными Саудовскому центральному банку, выпущенные 30 июня 2021 года и обновлённые вторым изданием 8 мая 2024-го, устанавливают действующие роли совета, исполнительного руководства, комитетов, контроля и управления рисками. Это

не специальная норма о подписании решений моделей ИИ. Применение требует перевести смысл принципов в проверяемые полномочия и процедуры, не выдумывая текстовых обязанностей.12

Совет саудовского банка в предлагаемой внутренней проверке мог бы спросить о доле решений, «принятых ИИ». Вопрос современный, но неполный. Точнее спросить: на каком этапе результат модели превращается в последствие, меняющее право или выгоду; кто утвердил этот этап; какой ущерб возможен до вмешательства уполномоченного сотрудника; может ли получатель просьбы об исправлении изменить результат или только отправить ещё одно письмо?

Если комитет по рискам одобрил модель, а продуктовая команда без должного разрешения изменила порог, тест точности сам по себе не найдёт дефекта в **правах решения (Decision Rights)**. Если поставщик может сменить производственную версию без подходящих договорных и технических ограничений, название банка на отказе не показывает, куда ушёл контроль. Если система учится на новых взаимодействиях, её поведение меняется, пока подпись на прежней версии спокойно лежит в папке соответствия, описывая исчезнувший мир.

Исследование Института финансовой стабильности Банка международных расчётов (Financial Stability Institute, BIS) от декабря 2024 года относит к вызовам ИИ в финансах управление, контроль моделей и зависимость от третьих сторон. Оно не оценивает каждый саудовский проект, но показывает: разделение знания между банком и поставщиком — не мелкий местный вопрос.13

Действенное полномочие исправления: предлагаемое понятие, а не готовый сертификат

Предлагаем здесь **действенное полномочие исправления (Executable Correction Authority)** как **аналитический инструмент статьи**, а не признанную теорию или действующий нормативный стандарт. Для решения с последствиями оно соединяет три условия: названного ответственного за проверку; достаточную информацию об основании и пределах; практическое и правовое полномочие изменить или остановить результат, пока исправление ещё имеет смысл.

Проверить идею можно без изобретения священного показателя. Организация выбирает заявленную выборку реальных значимых решений, безопасно воспроизводит сценарий законного возражения и спрашивает: удалось ли связаться с уполномоченным; смог ли он восстановить существенные причины без догадок; смог

ли при доказанной ошибке исполнить исправление в объявленный срок? Следует сообщать долю случаев, прошедших **все три условия**, с разбивкой по виду и тяжести, а не среднее, скрывающее худший вред.

Показатель **не доказывает правильность исходного решения**, не создаёт права отменить необратимое и допускает улучшение за счёт лёгкой выборки или щедрого срока. Он выявляет один аспект подотчётности — способность воплотить исправление. Проверяемая гипотеза статьи: системы, рано определяющие это право и регулярно его испытывающие, будут реже сохранять **ошибочное решение после обнаружения ошибки**, чем системы с номинальной жалобой. До полевого контролируемого исследования это не научный результат.

Четыре вопроса честнее, чем «есть ли человек в контуре?». Видит ли уполномоченный решающие данные? Может ли оспаривать *политику*, а не только *результат*? Останавливает ли срочное решение до труднообратимого эффекта? Понимает ли выгодоприобретатель путь возражения без внутреннего языка моделей? Совместное руководство ICO и Института Алана Тьюринга (The Alan Turing Institute) призывает объяснять затронутым людям решения с ИИ; объяснение — часть институциональных отношений, а не показ одной математики.14

Сильнейшее возражение: люди не справедливее только потому, что они люди

Не обязательно возвращать каждый файл сотруднику с его опытом и интуицией. Люди непоследовательны и устают; суждения меняются с настроением, опытом и нагрузкой. Хорошо испытанная модель с едиными критериями может уменьшить некоторые различия и найти пропущенные ручной проверкой ошибки. Рассмотрение каждой операции человеком само причиняет вред задержкой. Это серьёзный довод.

И случай 2020 года не доказывает непригодность всех статистических решений. Управление регулирования статистики признало, что сотрудники добросовестно стремились к точным результатам при чрезвычайных ограничениях времени и данных, а общий отказ от статистических моделей лишил бы общество реальной пользы.7 Человеческое решение тоже может быть необъяснимым, если причины не документированы.

Возражение уточняет применение, а не опровергает тезис. Подотчётность не требует повторять каждый расчёт человеком. Надзор распределяется по риску: ясная автоматизация простых обратимых случаев; проверка чувствительных границ;

обязательная остановка за пределами делегирования; независимая выборочная проверка, включая случаи, которые сотрудник одобрил без изменения. Это не единый рецепт для всех отраслей, а способ соразмерить власть и последствие.

Руководство ICO по управлению ИИ требует знать, кто определяет цели и средства обработки данных и кто обрабатывает их для другого; правовая роль меняется по этапу и отношению. Нельзя исчерпать ответственность договором, назвав поставщика «обработчиком», если фактический контроль и юридическая компетенция иные. Это норма защиты данных в своём диапазоне, а не общий вердикт о гражданской или банковской ответственности за все результаты ИИ.15

Решающая проверка — не чей палец касается кнопки, а кто может сказать «это правило неприемлемо» и вызвать действенное институциональное последствие. Без такой способности множество подписей не спасёт систему. В худшем случае подпись становится обрядом распределения вины после того, как власть уже распределили.

До следующего результата

Англия переиздала оценки того лета на другом основании. Нельзя было вернуть учеников в экзаменационные залы 13 августа, а перерасчёт не возвращал дни, прожитые под действием первого результата. Отмена была одновременно важной и ограниченной. Решение — не просто число в базе: оно входит в жизнь людей в конкретный час и продолжает движение после исправления.

Глубокий урок Ofqual — не простое предупреждение о «несправедливом алгоритме». Суждение бывает ясным на бумаге и мутным на карте полномочий. По мере того как модели пишут рекомендации, определяют риски и отправляют решения, процессы всё реже будут останавливаться ради явного человеческого мнения. Председателю совета следует спрашивать не «какая модель решила?», а «какое моё решение позволило модели создать это последствие и кто остановит его, если мы ошиблись?».

В напечатанной в августе 2020 года оценке было мало букв. За каждой стояла длинная цепочка решений. Современная организация, усвоившая урок, не ограничится подписью в конце цепи. Она найдёт того, у кого перо было в её начале.

Список литературы

1. Awarding GCSE, AS & A levels in summer 2020: interim report, Ofqual, 13 августа 2020 года; первичный промежуточный регуляторный отчёт о механизме, контексте и ограничениях в Англии.

2. Requirements for the calculation of results in summer 2020, Ofqual, выпущено 7 июля и изменялось до 20 августа 2020 года; первичные требования, страница содержит версии после смены политики.
3. Grading gaps in summer 2020, Ofqual; последующий официальный анализ 457 420 предметных записей 246 110 учеников заданного возраста и территории, не измерение истинности каждой оценки и не всё население Великобритании.
4. Written statement from Chair of Ofqual to the Education Select Committee, Roger Taylor, Ofqual, 2 сентября 2020 года; первичное заявление о министерском указании и ответственности регулятора.
5. Separation of Ownership and Control, Eugene F. Fama и Michael C. Jensen, *Journal of Law and Economics*, 1 июня 1983 года; основополагающее различие управления и контроля решений, не эксперимент об ИИ.
6. Formal and Real Authority in Organizations, Philippe Aghion и Jean Tirole, *Journal of Political Economy*, 1997; теоретическое исследование формальной власти и фактического влияния при распределённом знании.
7. Ensuring statistical models command public confidence, Office for Statistics Regulation, 2 марта 2021 года; относительно независимый официальный обзор уроков оценивания 2020 года в странах Великобритании, не судебное решение.
8. 12 CFR §1002.9: Notifications, Consumer Financial Protection Bureau, действующая американская норма, проверенная 31 октября 2026 года; первичный текст о причинах неблагоприятных решений только в своём диапазоне.
9. Consumer Financial Protection Circular 2022-03, май/июнь 2022 года, историческое толкование; Withdrawn Guidance фиксирует отзыв этого и циркуляра 2023-03 12 мая 2025 года. Они не представлены как действующие указания.
10. AI Risk Management Framework 1.0: Core, GOVERN, NIST, 2023; добровольная рамка для ролей руководства и управления модельным риском.
11. Article 14: Human Oversight, Европейская комиссия, законодательный текст в сводной редакции 27 июля 2026 года; официальный график 2026 года — AI Act: Regulatory framework — различает вступление и поэтапное применение.
12. Key Principles of Governance in Financial Institutions Subject to Saudi Central Bank's Oversight and Supervision, Саудовский центральный банк, 30 июня 2021 года, второе издание 8 мая 2024 года, помечено как действующее; первичный источник финансового управления, не специальная норма подписания решений ИИ.
13. Regulating AI in the financial sector: recent developments and main challenges, Financial Stability Institute/BIS, 12 декабря 2024 года; сравнительное институциональное исследование, не измерение конкретных саудовских банков.
14. Explaining decisions made with AI, ICO и Alan Turing Institute через GOV.UK, 27 января 2025 года; практическое руководство по объяснению, не полевой эксперимент.
15. What are the accountability and governance implications of AI?, ICO, официальное руководство на дату проверки; объясняет роли контролёра и обработчика в защите данных, а не все формы юридической ответственности.