SAFWAN ALMODHAYAN



АВТОРСКАЯ АНАЛИТИЧЕСКАЯ СТАТЬЯ

Редакционная дата: 1 ноября 2026 г.

Предсказание, изменившее предсказанное им будущее

> Раньше сигнал предупреждал нас о том, что может произойти. Затем мы решили действовать на его основе. А когда событие не случилось, потребовали от сигнала доказать свою правоту — словно судили пожарного за то, что дом, в который он вошёл, не сгорел.

Больничные отделения, услышавшие сигнал, которого не слышали другие

С октября 2019 по июль 2021 года пять больниц Медицинской службы Министерства национальной гвардии Саудовской Аравии провели эксперимент, смысл которого трудно свести к словам «технология сработала». Сорок пять терапевтических и хирургических отделений в девять этапов переходили от обычной помощи к электронной системе раннего выявления сепсиса — опасной патологической реакции организма, способной сопровождать инфекцию. Это не был робот, тайно ставящий диагноз, и не языковая модель, сочиняющая лечебные рекомендации. Система была проще — и потому полезнее для нашего вопроса.

Электронная медицинская карта рассчитывала сигнал по трём показателям быстрой последовательной оценки органной недостаточности (quick Sequential Organ Failure Assessment, qSOFA): снижению систолического давления до заданного уровня, повышенной частоте дыхания или изменению сознания. Когда в установленном

временном окне появлялись два признака, система выдавала предупреждение. В отделениях, перешедших к вмешательству, сигналы показывались медицинской команде и запускали маршрут реагирования; в остальных для сравнения использовался скрытый режим. Это позволило задать вопрос важнее точности расчёта: что меняется, когда у людей появляется сигнал, на который можно действовать?1

В исследование вошли 60 055 пациентов: 29 442 в группе электронного скрининга и 30 613 в группе сравнения. После предупреждения в группе вмешательства чаще назначали анализ на лактат и внутривенное введение жидкостей. Но таблица смертности без контекста озадачивает: необработанная доля внутрибольничных смертей в течение девяноста дней составляла **3,2%** при вмешательстве и **3,1%** при сравнении; явного преимущества скрининга в сыром показателе не было. После предусмотренной исследованием поправки на временные периоды, кластеризацию пациентов внутри отделений и больниц и статус COVID-19 скорректированный относительный риск оценили в **0,85**, с 95-процентным доверительным интервалом от 0,77 до 0,93. Это оценка относительного снижения риска после корректировки, а не утверждение, будто каждый выживший пациент спасён программой или будто одного сырого сравнения достаточно.1

Почему сырая и скорректированная оценки расходятся? Эксперимент проводился с поэтапным переходом: отделения меняли режим в разные сроки, затем пандемия COVID-19 изменила состав пациентов и риски в разные периоды. Авторы указали, что оценка эффекта требовала учитывать эти обстоятельства. Для исследования предсказаний это принципиально: даже рандомизация не выдаёт причинный ответ одной строкой чисел. Нужно знать, когда и как группы подвергались вмешательству и как менялась окружающая среда.1

Это не доказательство того, что модель ИИ предсказала риск и потеряла точность после спасения пациентов. Саудовская система использовала электронное правило qSOFA; статья не проверяла гипотезу о дрейфе производительности модели машинного обучения после внедрения. Но это экспериментальное подтверждение петли, о которой мы говорим: **сигнал риска меняет действия персонала, а действия персонала могут изменить событие, которое мы пытаемся предсказать**. Отсюда начинается проблема, совсем не похожая на качество классификации.

Что значит, когда успешный сигнал ошибается?

Представим пациента с признаками ухудшения. Система риска оценивает высокую вероятность тяжёлого события в следующие двадцать четыре часа. Предупреждение приходит команде; пациента быстро осматривают, проводят нужное вмешательство, и

событие не наступает. Через месяц отдел данных возвращается к записи: «Система предсказала опасность, но её не было. Разве это не ложное срабатывание?»

Возможно. А возможно, и нет.

Если оценка модели означала «вероятность события **при продолжении обычной помощи без нового предупреждения**», нельзя проверять её так, словно она означала «вероятность события **после нашей реакции на предупреждение**». Новая информация изменила путь, окончание которого требовалось предсказать. Если же модель с самого начала предназначена для прогноза после ожидаемой реакции команды, сравнение с фактическим исходом при такой политике правомерно — но это уже другая цель.

В статистике риск события при обычной политике можно записать как $R_0(x) = P(Y(0) = 1 \mid X = x)$, а риск при программе предупреждения и реагирования — как $R_1(x) = P(Y(1) = 1 \mid X = x)$. Здесь X — сведения, известные до действия; $Y(0)$ — потенциальный исход при **обычной помощи без новой программы**, а не в мире без медицины и лечения; $Y(1)$ — потенциальный исход при программе предупреждения. Разница отражает то, что политика может изменить у сходных людей при заданных допущениях дизайна.

Для одного пациента оба исхода одновременно не наблюдаются. Мы видим то, что произошло на выбранном пути, но не вторую версию мира с тем же пациентом, где команда не услышала предупреждения. Это называют проблемой потенциальных исходов (Potential Outcomes). Поэтому фраза «ничего не произошло» менее содержательна, чем кажется. Ничего не произошло **после чего**? После какого вмешательства? Что, вероятнее всего, случилось бы при другом пути?

В 2020 году Хуан Пердомо (Juan Perdomo), Тиана Зрнич (Tijana Zrnic), Селестин Мендлер-Дюннер (Celestine Mendler-Dünner) и Мориц Хардт (Moritz Hardt) предложили рамку «перформативного предсказания» (Performative Prediction). Если модель поддерживает решение, она может изменить будущее распределение данных, по которому её затем станут судить: мир после публикации прогноза уже не совпадает с миром, который она прогнозировала. Работа изучала математические свойства устойчивости при повторном обучении и взаимного влияния прогноза и данных; она не удостоверяет, что всякий сигнал, чья точность снизилась, успешно предотвратил событие.2

Разница между пророчеством, меняющим наше поведение, и пророчеством, которое ничего не меняет, не философская мелочь. Прогноз дождя обычно не меняет облака; прогноз ухудшения состояния пациента может изменить решение врача, прогноз школьного отставания — способ поддержки ученика, прогноз промышленной аварии — график обслуживания. Здесь ИИ уже не зеркало будущего, а новая переменная в его создании.

Больница, нашедшая сигналы риска в действиях медсестёр

Нужен пример с настоящим машинным обучением, а не только правилом предупреждения. В апреле 2025 года *Nature Medicine* опубликовала рандомизированное испытание системы раннего предупреждения об ухудшении состояния **CONCERN**. Она опирается не только на анализы: система учится на временных закономерностях сестринской документации и наблюдения. Когда медсестра начинает чаще измерять показатели? Когда оставляет необычную запись? Когда проводит наблюдение в нетипичное время? Изменение поведения команды становится сигналом, опережающим некоторые изменения в организме пациента.3

Испытание охватило **74 клинических подразделения**, случайно распределённых между вмешательством и обычной помощью в двух американских системах здравоохранения, и **60 893 госпитализации**. Отображение сигналов CONCERN в рабочем процессе связывалось со снижением скорректированного мгновенного риска внутрибольничной смерти: отношение рисков **0,64**, 95-процентный доверительный интервал от 0,53 до 0,78. Показатель 0,64 не означает, что «64% пациентов выздоровели благодаря ИИ»: это отношение сравнительных мгновенных интенсивностей риска в анализе выживаемости. В представленном авторами анализе сократилась и продолжительность пребывания.3

Но в той же таблице скорректированный мгновенный риск **неожиданного перевода в отделение интенсивной терапии** вырос: отношение рисков **1,25**, доверительный интервал от 1,09 до 1,43. Как смертность может снижаться, а показатель, похожий на ухудшение, — расти? Возможно, раннее выявление побуждает чаще переводить пациентов на более высокий уровень помощи, но исследование не позволяет признать каждый перевод успешным спасением. В иных случаях неожиданный перевод может быть следствием ухудшения, которого следовало избежать. Нужны время, причина, клиническая уместность и изменение исходов, а не знак плюс или минус перед числом.3

Управленческий отчёт здесь становится двуликим. Один отдел скажет, что система улучшила опасный исход; другой заметит рост неожиданных переводов и потребует объяснения. Оба читают истинный результат. Но их расхождение выявляет более важное: **успешный сигнал может изменить карту событий, и нельзя объявлять любое новое событие успехом или неудачей, пока не понят путь вмешательства, который его породил**.

Авторы указали ограничения: исследование прошло в двух американских системах здравоохранения в период, затронутый COVID-19; различались наборы отделений, практики документирования, а на некоторые результаты могли влиять неизмеренные

факторы. Поэтому его вывод сильнее маркетинговой истории, но уже универсальной гарантии для каждой больницы и системы предупреждения.3

В школе тоже: сигналы риска снизились, итоговый показатель не сдвинулся

Чтобы не ограничивать парадокс реанимацией, перенесёмся в американские старшие школы. В 2014/15 учебном году исследование, позднее опубликованное Министерством образования США, испытало систему раннего предупреждения и вмешательства **EWIMS** для учеников с риском не окончить школу. **73 школы** распределили случайно: 37 внедрили программу, 36 сохранили обычные процедуры на год оценки. В исследование вошёл **37 671 ученик** первых двух классов старшей школы. Это была не языковая модель и не современная система машинного обучения, а комплекс ранних индикаторов, наблюдения и вмешательства.4

Через год доля хронических прогулов составила **10%** в школах программы и **14%** в группе сравнения. Хотя бы один предмет не сдали **21%** учеников против **26%**; оба различия были статистически значимыми. Но доля учеников, не набравших достаточно учебных часов для выпуска за четыре года, составила **14% в обеих группах**: выявляемого эффекта на этот показатель учебного продвижения за один год не было. Авторы также отметили трудности с выполнением семи шагов программы и в целом слабую реализацию во многих школах.4

Урок в том, что поведение, создающее риск, может измениться раньше конечного желаемого исхода. Хронические пропуски — сигнал, провал предмета — сигнал, продвижение к выпуску — результат ближе к дальней цели. Если обучить новую систему на записях после вмешательств, предотвративших часть прогулов или неуспеваемости, а каждого улучшившегося ученика считать «ошибочной классификацией риска», мы смешаем вопросы «Кому была нужна поддержка?» и «Кто не справился, несмотря на поддержку?» Но если посещаемость улучшилась, а продвижение к выпуску — нет, объявлять конечную цель достигнутой тоже нельзя.

Школа и больница различаются технологией и определением успеха. Общее у них одно: когда знак риска используется для изменения человеческого поведения, последующие результаты становятся частью **реакции учреждения**, а не нейтральным экзаменом прогностической силы знака.

Недостаточно выбрать человека с высоким риском

Это различие необходимо, чтобы «скрытый успех сигнала» не превратился в удобное оправдание любой слабой модели. В Камдене, штат Нью-Джерси, испытали интенсивную программу координации помощи для пациентов с частыми госпитализациями и сложными потребностями. Рабочая гипотеза была привлекательна: мы знаем, кто постоянно возвращается в больницу; предоставим этим людям согласованную медицинскую и социальную поддержку, и повторных госпитализаций станет меньше.

Однако рандомизированное испытание с **800 пациентами** не выявило статистически значимого снижения повторных госпитализаций за 180 дней. Показатель составил **62,3%** в программе и **61,7%** при обычной помощи; скорректированная разница — **0,82 процентного пункта**, доверительный интервал от снижения на 5,97 до роста на 7,61 пункта. Особенно важно, что сравнение участников программы с самими собой до и после вмешательства создало бы впечатление резкого падения примерно на **38 процентных пунктов**. Они вошли в исследование после периода исключительно интенсивного пользования помощью, а затем число госпитализаций снизилось в обеих группах. Автоматически приписывать это программе нельзя.5

Камденская программа не была опытом с ИИ. Но она даёт решающее экспериментальное возражение против смешения **поиска людей с наибольшей вероятностью события** и **поиска людей, у которых доступное вмешательство способно предотвратить событие**. Человек с самым высоким риском не обязательно получит наибольшую ожидаемую пользу от конкретной программы. Причины риска могут не отвечать на это вмешательство, требовать отсутствующих ресурсов или по природе быть временными.

На бумаге две модели могут одинаково ранжировать пациентов по вероятности повторной госпитализации и согласиться в том, кто находится в наибольшей опасности. Но вопрос о том, кому больше поможет определённое вмешательство, может потребовать совершенно другого порядка. Один вопрос касается **уровня риска**, другой — **изменяемости риска**. Их смешение придаёт высокой прогностической точности видимость превосходной профилактической политики без доказательств.

Та же трудность проявилась в опубликованном в феврале 2026 года исследовании пассивного отображения прогнозов ИИ об ухудшении состояния пациентов кардиологического отделения. Оно охватило **10 422 госпитализации** и показывало траектории риска, не предписывая определённой клинической реакции. Улучшения основного исхода не обнаружили. Врачи переместили около **11%** пациентов между

койками с отображением и без него, ослабив некоторые свойства практической рандомизации. Это не приговор всем моделям предупреждения, а напоминание: **знание риска не создаёт автоматически действие, способное этот риск снизить.**[6]

Итак, обе возможности нужно держать вместе: сигнал мог быть верен, а вмешательство — спасти пациента; но сигнал мог быть неточен или вмешательство бесполезно. Различает варианты причинный дизайн, а не доброжелательное отношение к алгоритму.

Нужный нам исход — не тот исход, который у нас есть

В статье 2020 года Нан ван Геловен (Nan van Geloven) и соавторы подчеркнули: нужно определить, что именно предсказывает модель, если лечение начинается после измерения риска. Такое определение они назвали **предиктимандом** (Predictimand) по аналогии с целевой величиной эффекта в клинических испытаниях. Предсказываем ли мы исход без начала дополнительного лечения? Исход при продолжении лечебной практики из обучающих данных? Или событие до начала лечения? Это разные вопросы с разными ответами, методами оценки и допущениями.[7]

В октябре 2025 года Кристофер Бойер (Christopher B. Boyer) и соавторы напрямую рассмотрели **построение и оценку контрфактических прогностических моделей**. Они объяснили, почему качество модели исхода при гипотетической лечебной политике нельзя оценивать только по данным, наблюдавшимся при другой политике, и предложили методы оценки потерь, калибровки и площади под ROC-кривой в контрфактическом мире с явными условиями статистической идентификации. Примером служила модель риска сердечно-сосудистых заболеваний у людей, не начавших лечение статинами. Это условные методологические результаты, а не рецепт, делающий ненаблюдавшиеся исходы видимыми.[8]

Нужно различать три величины:

1. **Исходный риск:** вероятность вреда при эталонной политике, например обычной помощи без новой системы предупреждения.
2. **Риск после вмешательства:** вероятность вреда при применении сигнала по известной политике и фактической реакции людей.
3. **Польза политики:** величина вреда, устранённого переходом от первой политики ко второй с учётом побочных эффектов, стоимости, ресурсов и иных последствий.

Исходный риск может быть высоким, риск после вмешательства — низким, а польза политики — большой. Исходный риск может быть столь же высоким, но бесполезное вмешательство почти не создаст пользы. Для некоторых групп вмешательство способно

причинить вред. Одной «оценки риска» недостаточно, чтобы выбрать между вариантами.

Только для числовой иллюстрации представим человека А с риском события 30% при обычной политике, который снижается до 10% при определённом вмешательстве, и человека Б с риском 20%, снижающимся до 19%. Оценочная разница пользы — 20 процентных пунктов для первого и один пункт для второго. Это **гипотетические числа для объяснения, а не результаты исследования**; наблюдение единственного исхода человека не раскрывает его индивидуальную реализованную пользу. Мы можем лишь пытаться оценить средние эффекты для групп при подходящих допущениях и дизайне.

Проблема усиливается, когда система периодически переобучается на данных, созданных её собственным применением. Увидев, что люди с высокими оценками не пережили событие после лечения, перевода или поддержки, модель усваивает новую связь между их признаками и наблюдаемым исходом. Она может лучше предсказывать то, что произойдёт **при нынешней политике**, и хуже оценивать то, что случилось бы **без неё**. Переобучение не всегда ошибка, но определение вопроса должно оставаться устойчивым. Нельзя молча заменить предсказываемую величину и назвать это обычным обновлением версии.

Почему точность вводит нас в заблуждение

Высокая площадь под ROC-кривой может подтверждать способность модели различать исходы в конкретной тестовой выборке, но сама по себе не отвечает на вопрос о влиянии применения на пациентов, учеников или организации. Если после успешного вмешательства прогностическая метрика снижается, её нельзя толковать без знания исходной цели и политики. Распределение событий могло измениться из-за вмешательства, состав населения — сдвинуться, а качество модели — действительно ухудшиться.

Есть три разных суда, и ни один не должен поглощать два других:

Первый — суд над прогнозом: были ли оценки риска корректны в заданном эталонном мире? Сюда входят калибровка (Calibration), различающая способность (Discrimination), качество данных и соответствие периода и популяции.

Второй — суд над реакцией: изменило ли сообщение поведение принимающего решение? Сколько времени прошло от сигнала до осмотра или действия? Было ли действие уместно и откликнулась ли на него целевая группа?

Третий — суд над исходом: каков эффект политики по сравнению с обычной помощью или альтернативой? Уменьшился ли вред? Возникли ли другие виды вреда, расходы, ненужные переводы или истощение работников и ресурсов?

Эти суды нельзя сжать в один процент. В саудовском испытании рост числа анализов и назначений жидкости был частью реакции персонала, а смертность — отдельным исходом. В CONCERN один риск снизился, а переводы в интенсивную терапию выросли. В школах EWIMS улучшились некоторые индикаторы без выявленного годичного эффекта на продвижение к выпуску. В Камдене программа не доказала пользу по основной цели, хотя была направлена на людей с интенсивным пользованием помощью. Механизмы различны, но научная оценка всегда требует отделять прогноз, действие и эффект.

Здесь проблема не та же, что при отсутствии меток у людей, которым не дали возможности проявить исход. В профилактическом прогнозировании результат может быть **полностью наблюдаемым**, но он произошёл **после вмешательства политики**. Отсутствие события в записи само по себе не говорит, возникло бы оно без вмешательства. Это не пустое поле в таблице, а отсутствующая альтернативная траектория мира.

Как испытывать модель, предотвращающую часть предсказанных событий?

Когда это этически и организационно допустимо, существует прямой экспериментальный путь: сохранить обычную помощь или услугу для всех и случайно распределить **добавление новой программы** между сходными подразделениями либо вводить её по продуманным этапам. Прогнозы записываются до показа, фиксируются время показа, действия работников и последующие исходы. Сравнение не означает отказа от обязательного лечения или сознательного подвергания человека известному риску. Проверяется дополнительная функция, польза которой ещё не доказана. В чувствительных областях дизайн должны одобрить профильные специалисты и органы этического управления.

Некоторые системы позволяют рассчитывать риск в скрытом режиме на время сравнения при защите данных и соблюдении границ обычной помощи. Эффект **доступности предупреждения** измеряется на уровне групп или отделений, без притязания на знание того, кто именно из выживших умер бы. Поэтапные и кластерные рандомизированные испытания полезны, когда знания распространяются между работниками одного подразделения, но им свойственны сложности периодов и общих событий — как показал COVID-19, сопровождавший саудовское больничное исследование.

Если рандомизация невозможна, можно попытаться эмулировать целевое испытание (Target Trial Emulation) на наблюдательных данных либо использовать методы учёта лечебного вмешательства и контрфактической оценки эффекта. Но они зависят от допущений, которые нельзя скрывать: известны ли вмешивающиеся факторы, влияющие и на решение, и на исход; можно ли сравнивать людей с реальной вероятностью разных вариантов; определено ли вмешательство последовательно; не искажено ли сравнение неучтённым взаимным влиянием между подразделениями. Не во всякой среде эти допущения выполняются.

Если вмешательство одному пациенту занимает койку интенсивной терапии, необходимую другому, дело не ограничивается изолированным индивидуальным эффектом. Возникают распределение ресурсов, перегрузка и взаимное влияние (Interference). Политика может помогать десяткам пациентов и давать иной итог при внедрении во всех отделениях страны. Поэтому недостаточно проверить точность модели и предположить, что удвоение числа сигналов удвоит число спасений. Возможности реагирования различаются, чрезмерные предупреждения обходятся дорого, а поведение команд меняется со временем.

Реестр прогнозов и вмешательств: опровержимое предложение

Я предлагаю хранить «**реестр прогнозов и вмешательств**» (Forecast–Intervention Ledger) — редакционную идею, а не утверждённый стандарт или систему с доказанной эффективностью. Он нужен не для умножения полей панели, а для разведения трёх вещей, прежде смешивавшихся: что было известно до вмешательства, что сделали из-за этого знания и что произошло после.

Для каждого подходящего случая фиксируются время сбора предшествующей информации, версия модели, определение риска, выраженного оценкой, горизонт прогноза, время сигнала, факт его показа принимающему решение, последовавшая реакция и её время, исходы и возможный вред. Отдельный аналитический уровень оценивает причинный эффект в сопоставимых группах и хранит неопределённость и допущения. Поле «что случилось бы с этим пациентом без предупреждения» не заполняют вымышленным индивидуальным числом: такое будущее не наблюдалось.

Реестр можно превратить в три раздельно рассматриваемые панели:

- **Панель прогноза:** калибровка моделей при эталонной политике, чувствительность сигнала и соответствие оценки риска целевой группе.
- **Панель вмешательства:** доля сигналов, вызвавших уместную реакцию, временные интервалы, нагрузка, направления и вред.

- **Панель эффекта:** разница исходов относительно определённой политики сравнения с доверительными интервалами, стоимостью и влиянием распределения ресурсов, а не просто наблюдавшееся у получивших сигнал.

Опровергающий ценность системы опыт должен быть понятен: если хорошее сравнение не показывает значимой разницы в важном исходе, вред предупреждений превышает пользу или выявление группы наибольшего риска не меняет полезность вмешательства, риторика «мы предотвратили невидимые события» систему не спасает. Модель должна выдерживать возможность, что она не улучшила мир, так же как мы допускали её успех до появления результатов.

Главное правило реестра: ответ никогда не звучит так — «после установки системы случаев стало меньше, значит, она их предотвратила». Число случаев могло снизиться из-за изменения населения, сезона, помощи или окончания предыдущего пика. Камденский урок особенно суров: сравнение до и после программы создавало видимость большого улучшения, но случайно выбранная группа сравнения прошла сходную траекторию.

Совету директоров больницы, министерству образования или органу общественной безопасности важны не число выданных сигналов и не только число дней без катастрофы. Важен вопрос: **насколько именно эта политика изменила вред, для кого, какой ценой и с какой уверенностью в выводе?**

Когда второе будущее становится частью научной ответственности

От ИИ не требуется знать сокрытое, а от исследователя — показывать на экране не случившееся будущее. Требуется признать: профилактические решения создают парадокс, который не разрешить одними традиционными тестами классификации. Одни благополучные исходы могут быть плодом успешного вмешательства, другие — чрезмерного сигнала, удачи или естественного изменения. Алгоритм не способен различить их, выбрав наиболее лестный рассказ о себе.

Поэтому нельзя вознаграждать модель за страх перед всем и нельзя наказывать лишь потому, что опасность не осуществилась. Нельзя смешивать риск пациента с его излечимостью, отставание ученика — с вероятностью улучшения при помощи учителя, а любое снижение показателя после программы — с эффектом программы.

Самая глубокая перемена, которую перформативное предсказание требует от организации, — переход от кажущегося научным вопроса «Предсказала ли система то, что произошло?» к более трудному и справедливому: «Что произошло бы, если бы мы

не действовали по её прогнозу, и как проверить разницу, не утверждая, будто мы видели оба будущих?»

Тогда меняется смысл успеха. Лучшая система предупреждения — не та, чьи предостережения чаще всего сбываются: осуществление риска не является целью профилактики. Но не обязательно и та, после которой исчезают все события: исчезновение могло быть не её заслугой. Лучшей окажется система, о которой эксперимент или дисциплинированный причинный вывод доказывает, что она помогла людям сделать плохой исход менее вероятным, не скрывая цены и альтернативного вреда.

Мы научились проверять, знает ли компьютер что-то о будущем. Осталось научиться проверять, **сделало ли его знание будущее лучше**.

Список литературы

1. Arabi и соавторы, JAMA, электронный скрининг сепсиса — печатная публикация 2025 года, онлайн-публикация 2024 года.
2. Perdomo и соавторы, Performative Prediction, ICML 2020.
3. Rossetti и соавторы, рандомизированное испытание CONCERN, Nature Medicine, 2025.
4. Institute of Education Sciences, исследование воздействия EWIMS, апрель 2017 года.
5. Finkelstein и соавторы, рандомизированное испытание интенсивной координации медицинской помощи, NEJM, 2020.
6. Keim-Malpass и соавторы, рандомизированное испытание аналитики на основе ИИ, Scientific Reports, 2026.
7. van Geloven и соавторы, Predictimands, European Journal of Epidemiology, 2020.
8. Boyer и соавторы, оценка контрфактических прогнозов, Statistics in Medicine, 2025.