

Когортирование на базе ИИ незаметно меняет способы формирования доказательств на основе реальных данных (RWE)

Источник: MedCity News

Оригинал: <https://medcitynews.com/2026/04/ai-powered-cohorting-is-quietly-reshaping-how-real-world-evidence-gets-built/>

RWE

аналитика данных

искусственный интеллект

клиническая разработка

семантическое картирование

электронные медкарты

Формирование когорт пациентов долгое время было одним из наиболее значимых и наименее заметных этапов в аналитике здравоохранения. Независимо от того, является ли целью получение доказательств из реальной клинической практики (**RWE** — Real-World Evidence), исследования экономики здравоохранения и исходов (**HEOR** — Health Economics and Outcomes Research), клиническая разработка или коммерческое планирование, почти каждый последующий аналитический вывод зависит от обманчиво простого действия: определения того, кто именно подлежит учету.

Создание когорт редко обсуждается за пределами технических групп. Его часто рассматривают как механическое упражнение по составлению запросов, а не как то, чем оно является на самом деле: процесс преобразования клинического смысла в вычислимую логику, выполняемую на основе фрагментированных, несовершенных данных.

По мере того как искусственный интеллект внедряется в аналитику здравоохранения, процесс формирования когорт претерпевает тихую, но важную трансформацию. Эта трансформация касается не только

разговорных интерфейсов или ускоренной генерации запросов. Речь идет о перестройке архитектуры, лежащей в основе самого процесса создания когорт. Сейчас намечается более структурный сдвиг: рабочие процессы на базе ИИ, которые распознают формирование когорт как многомерную проблему, охватывающую клинические нюансы, вариативность кодирования и временную логику, и, следовательно, рассматривают этот процесс как последовательность явных, валидированных шагов, а не как единую непрозрачную операцию.

Формирование когорт — это не одна проблема, а цепь взаимозависимых задач, где каждый шаг зависит от целостности предыдущего. Важное определение когорты должно правильно интерпретировать клинические концепции, применять критерии включения и исключения, учитывать временные аспекты и работать с «грязными» данными из реальной практики. Если идентификация пациента нестабильна, ломается лонгитюдная (продольная) логика. Если клинические концепции отображены некорректно, критерии включения смещаются. Если процесс выполнения непрозрачен, валидация превращается в гадание.

Данные из реальной практики усложняют задачу. Данные страховых требований (claims) и электронных медицинских карт (**EHR** — Electronic Health Record) по своей сути неполны. Пути пациентов фрагментированы между различными плательщиками и системами, а правила кодирования различаются. Незначительные несоответствия, такие как неправильно примененное окно включения или неверно интерпретированный класс лекарственных средств, могут существенно изменить количество пациентов или внести систематическую ошибку в последующий анализ.

Данные здравоохранения представляют еще одну проблему: они закодированы, а не семантически. Диагнозы, процедуры, лекарства, лабораторные исследования и визиты представлены через такие словари, как **ICD-10**, **CPT**, **HCPCS**, **NDC** и **LOINC**. Эти системы были разработаны для выставления счетов и документирования, а не для аналитической ясности.

То, что кажется простым клиническим понятием — например, пациенты с диабетом 2 типа, принимающие метформин, и имеющие как минимум два измерения ИМТ (индекса массы тела) выше 35 за последние два года — быстро разрастается до сотен отдельных кодов в словарях диагнозов, процедур, лабораторных исследований и аптечных справочников. Семантическое картирование — это процесс перевода этого клинического намерения в обоснованную, исполняемую логику во всех этих словарях.

Традиционные подходы справляются с этой сложностью за счет ручной тщательности: тщательно написанных SQL-запросов, ручного поиска кодов, итеративной валидации и экспертной проверки. Хотя эта модель эффективна, она не масштабируется. Она ограничивает количество вопросов, которые организация может обоснованно задать, и скорость, с которой она может адаптироваться к новым доказательствам.

Почему однослойные подходы не справляются

Многие аналитические инструменты с поддержкой ИИ пытаются упростить формирование когорт, используя одну большую языковую модель для перевода естественного языка в исполняемую логику. Хотя это может ускорить базовые запросы, модель сталкивается с трудностями, когда требуются точность, объяснимость и воспроизводимость.

От одной модели требуется одновременно интерпретировать намерение, применять клиническую логику, управлять временными связями, выполнять запросы, валидировать результаты и объяснять полученные данные. Когда возникают ошибки, трудно определить их источник или оценить их влияние. В средах, где важны воспроизводимость и возможность аудита, непрозрачность ограничивает доверие.

Альтернатива в виде рабочего процесса

Более надежный подход заключается в том, чтобы рассматривать формирование когорт как рабочий процесс (workflow), а не как единую задачу ИИ. В этой модели процесс разбивается на отдельные этапы, каждый из которых обрабатывается специализированным компонентом, предназначенным именно для этой цели.

Такие ориентированные на рабочий процесс архитектуры используют агентную структуру, в которой различные модели отвечают за интерпретацию намерений, разрешение клинических концепций, временные рассуждения, выполнение, валидацию и объяснение. Каждый шаг выдает немедленные результаты, которые можно проверить, просмотреть и уточнить.

Это отражает то, как уже работают опытные аналитики, но теперь с применением автоматизации на каждом этапе. Результатом становится более быстрая итерация без потери ясности относительно того, как была сформирована когорта или почему она могла измениться.

Одним из наиболее заметных эффектов формирования когорт на основе рабочих процессов является скорость. Определения, которые раньше требовали недель согласований, теперь можно исследовать за часы. Ввод данных на обычном языке можно уточнять итеративно, получая количество пациентов на каждом этапе.

Для команд, работающих в сферах **HEOR**, **RWE** или клинической разработки, это меняет экономику исследований. Вместо того чтобы отдавать приоритет небольшому количеству «безопасных» анализов, команды могут тестировать больше гипотез, изучать пограничные случаи и исследовать редкие подгруппы населения, что в противном случае было бы непрактично.

Важно отметить, что это ускорение не происходит за счет снижения стандартов. Системы на основе рабочих процессов полагаются на специализированные валидированные модели, а не на модели общего назначения. Результаты прозрачны: исполняемая логика и журналы аудита доступны для проверки. Валидация относительно эталонных наборов данных встроена в рабочий процесс. Именно этот акцент на воспроизводимости и объяснимости позволяет формированию когорт с помощью ИИ перейти от экспериментов к рутинному использованию.

Место запуска интеллекта имеет значение

Еще один важный сдвиг касается развертывания. Многим инструментам ИИ требуется перемещение данных в отдельные среды для доступа к расширенным функциям, что создает трудности в вопросах безопасности, управления и операционного доверия.

Альтернативным паттерном является развертывание аналитических сервисов внутри существующих сред данных. При таком подходе данные остаются на месте, а к ним подтягивается аналитическая логика. Эта модель более естественно соответствует требованиям медицинских организаций к конфиденциальности и управлению данными и снижает барьеры для внедрения. Это также позволяет формированию когорт на естественном языке функционировать как встроенная возможность, а не как отдельное приложение. Команды могут вызывать рабочие процессы через API, аналитические блокноты (notebooks) и многоагентные системы, интегрируя их в инструменты, которые они уже используют.

Организации, использующие рабочие процессы формирования когорт на базе ИИ, сообщают о существенном росте эффективности. Определения, которые когда-то требовали недель программирования и валидации, теперь могут

быть исследованы за минуты с явным указанием логики и допущений. В более широком смысле это указывает на сдвиг в восприятии формирования когорт. Вместо разового индивидуального упражнения определения когорт становятся многоразовыми активами. Логика можно стандартизировать, совершенствовать и передавать между командами. Со временем формирование когорт превращается в часть аналитической инфраструктуры организации, а не в повторяющееся техническое препятствие.

Глобальные последствия

Значимость формирования когорт с помощью ИИ заключается не столько в интерфейсе, сколько в рабочем процессе, спроектированном под ним. Разбивая сложные аналитические задачи на явные, валидированные шаги, организации могут двигаться быстрее, сохраняя при этом строгость методов.

В аналитике здравоохранения, где последующие решения так сильно зависят от того, кто учтен и почему, этот сдвиг может оказаться более значимым, чем многие более известные применения ИИ. Речь идет не о замене экспертов, а о кодировании их опыта в масштабируемые системы. Наибольшую выгоду от формирования когорт с помощью ИИ получают не просто те организации, которые движутся быстрее. Это будут те, кто внедрит мастерство управления данными пациентов и семантическую строгость в воспроизводимые, проверяемые рабочие процессы.

Поскольку доказательная база все чаще определяет как клинические, так и коммерческие решения, способность генерировать когорты, которые являются быстрыми, прозрачными и обоснованными, может стать одной из важнейших компетенций медицинских организаций. Но генерация когорт — это только отправная точка. Ценность заключается в изучении популяции, валидации определений, исследовании исходов и паттернов лечения, а также проведении последующего статистического анализа. Инсайты уточняются путем итераций: корректировки критериев, тестирования допущений и проверки результатов на прочность до тех пор, пока выводы не станут достаточно надежными, чтобы лечь в основу дизайна исследований, регуляторной стратегии, решений по доступу на рынок и клинической практики.

Глинн Деннис — главный научный сотрудник Kythera Labs. Он руководит инициативами компании в области наук о жизни и биофармацевтики, включая исследования и разработку продуктов. Глинн обладает более чем двадцатилетним опытом работы в области данных и ИИ в Kythera Labs, ранее

занимая многочисленные руководящие должности в научных областях, области данных и ИИ в различных компаниях отрасли, включая NIH, Genentech, Bio-Rad Laboratories и AstraZeneca.

Перевод выполнен: 13.04.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.