

Данные траекторий: как закрытый цикл «решение → исход» становится единственным реальным moat

2026-05-06

Данные траекторий: как закрытый цикл «решение → исход» становится единственным реальным moat

Что уже можно купить, а что — нет

К апрелю 2026 года стек вертикального AI разделился на то, что покупается из коробки, и то, что не покупается ни за какие деньги. За последний квартал четыре крупнейших провайдера — Anthropic, OpenAI, Google и AWS — независимо выпустили managed agent harness как отдельные продукты — этот сдвиг рынка businessengineer.ai назвал «harness as the agentic moat» ещё в марте, а к маю продуктизация уже завершилась. То, что год назад требовало месяца инженерной работы, ставится из SDK за несколько часов. Параллельно появилась индустрия памяти для агентов: MemO закрыл Series A на \$24 млн в октябре 2025 при оценке \$150 млн, Graphiti накопил десятки тысяч звёзд на GitHub, LongMemEval (arXiv:2410.10813) стал стандартным академическим бенчмарком долговременной памяти агентов.

Если модель меняется за выходные, а слой памяти переносится за 2–4 недели, вопрос звучит прямо: что именно нельзя купить, синтезировать или воспроизвести? Не инфраструктура памяти. А данные, которые через эту инфраструктуру прошли.

Конкретнее — это один тип данных: записи закрытого цикла. Каждая такая запись — формула вот что решил агент / человек, вот что произошло в итоге — это контракт между действием и его измеримым следствием через N дней. Если сформулировать точно, пара **решение → исход — это атомарная единица обучающего сигнала вертикального AI, состоящая из трёх связанных полей**: контекст, в котором было принято решение; само действие (что сделал агент или человек); измеримое последствие через заданный промежуток времени. Эта же сущность называется **trajectory data — структурированная пара контекста и фактического исхода, зафиксированная в момент реального действия и связанная во времени с последствием**. В отличие от обычного лога, такая запись содержит не только сам факт, но и связь действия с измеримым результатом. Это единственный слой защиты, который нельзя восстановить задним числом — сколько бы денег на восстановление ни потратить.

Три уровня, три разных горизонта жизни

Вертикальный AI-продукт сегодня строится на трёх слоях. У каждого из них — разный срок жизни как конкурентного преимущества.

Первый слой — модель. Горизонт защиты: 12–18 месяцев. Пока вы используете одну фронтинговую модель, конкурент использует другую; через год оба перейдут на следующее поколение, и разница исчезнет. Jason Lemkin из SaaSr формулирует это прямо: «AI is table stakes, prompts are portable, AI quality is not your moat» — то есть качество модели и переносимость промптов не создают устойчивого преимущества. Это знакомая логика технического преимущества, которое утекает вниз по стеку — её прошли реляционные базы данных, облачная инфраструктура, теперь проходят фронтинговые модели. На модели устойчивость не строится.

Второй слой — схема данных (representation layer). Горизонт защиты: годы. Это онтология домена: какие сущности существуют, как они связаны, что считается статусом, кто имеет право эскалировать, какие исключения валидны. Подробно про этот слой — в предыдущей заметке про representation layer как источник стоимости переключения. Чужому провайдеру нужны месяцы на корректное описание домена и ещё месяцы на калибровку схемы под реальное поведение конкретного рынка.

Третий слой — данные траекторий (trajectory data). Горизонт защиты: невозможно воссоздать ретроспективно. Это не «история действий». Это **закрытый цикл (closed loop) — замкнутая связь между решением в конкретном контексте и фактическим исходом, наблюдаемым через определённый промежуток времени.** Решение либо привело к нужному результату, либо нет. Прогноз сбылся или не сбылся. Рекомендация была принята клиентом или отвергнута. Именно эти пары «решение → исход» становятся обучающим сигналом, который улучшает следующие решения системы в похожих ситуациях. **Более умной моделью — это значит не модель с большим числом параметров, а система, у которой выше точность прогнозирования исхода в конкретном домене за счёт накопленных пар.** Их нельзя синтезировать, восстановить из логов или купить.

Почему «нельзя воссоздать» — не метафора, а физическое ограничение

Данные траекторий создаются только в момент реального решения в реальном контексте. Ни ретроспективная разметка логов, ни синтетические данные, ни дообучение на общедоступных датасетах не воспроизводят этот слой. Причина: запись содержит не только сам факт решения, но и контекст момента — состояние системы, историю предшествующих взаимодействий, конкретные условия, сложившиеся к тому часу. Этот контекст начинает теряться уже через сутки: параллельные процессы успевают изменить состояние, а исход решения проявляется позже самого решения, и связать одно с другим по логам постфактум становится невозможно (arXiv:2507.05257).

Исследования памяти агентов подтверждают асимметрию количественно. MemoryAgentBench, представленный в работе Hu, Wang и McAuley «Evaluating Memory in LLM Agents via Incremental Multi-Turn Interactions», вводит четыре ключевых компетенции для систем с памятью: точный поиск, обучение во время использования, понимание длинных горизонтов и избирательное забывание. Авторы фиксируют, что ни один из проверенных подходов — от простого расширения контекста до retrieval-augmented generation и внешних модулей памяти — не закрывает все четыре компетенции одновременно. На LongMemEval (arXiv:2410.10813) разрыв в категории temporal reasoning между системами с накопленной долговременной историей и без неё авторы оценивают в десятки процентных пунктов — и этот разрыв не закрывается увеличением контекстного окна.

MEM1 (arXiv:2506.15841) идёт дальше: работа показывает, что системы, обученные на реальных траекториях решений, формируют принципиально другие внутренние представления задачи — не описательные, а оперативные. Из этого следует, что разрыв в качестве между провайдером с историей и провайдером без неё не просто существует — он увеличивается с каждым месяцем работы в одном домене.

Gong: прецедент монетизации

Рынок уже научился продавать запись траекторий как отдельный продукт — за пределами AI-агентов. Самый показательный пример — Gong.io.

Gong записывает каждый разговор отдела продаж, извлекает из него паттерны и связывает их с исходами сделок. Это и есть запись траекторий в чистом виде — только для sales-команды, а не для AI-агента. Модель продукта состоит из трёх обязательных строк прайса: платформенная подписка (\$5 000–\$50 000/год), per-user лицензия (\$1 300–\$1 600 на пользователя в год) и единовременное внедрение (\$7 500–\$65 000). Для команды из 10 человек первый год обходится примерно в \$28 000, для 50 человек — около \$106 000.

Причина, по которой рынок готов платить эти деньги, проста: Gong не продаёт «аналитику звонков». Gong продаёт накопленную историю «вот что говорилось на этом этапе сделки — вот чем сделка закончилась». Платформенная подписка отделена от лицензии на пользователя именно потому, что хранилище и доступ к историческим записям — это базовая ценность, не зависящая от размера команды. Внедрение стоит отдельно и дорого по той же причине: клиент платит за то, чтобы правильно поставить запись траекторий с первого дня, а не пытаться разметить логи задним числом, когда контекст уже потерян.

Та же логика работает в LLM-observability. Langfuse продаёт запись траекторий агента от \$29 до \$2 499/мес в зависимости от глубины ретенции данных. Braintrust — от \$249/мес с отдельной строкой за хранение данных. Коммерческая логика общая: хранение и анализ траекторий — отдельная позиция прайс-листа, а не «инфраструктура внутри».

Вертикальный маховик против горизонтального «company brain»

К 2026 году оформилась отдельная категория горизонтальных «company brain» продуктов: корпоративная память для всей компании. Glean привлёк оценку в районе \$7 млрд в раунде 2025 года. Y Combinator включил enterprise memory layer в Request for Startups последних батчей. Эта категория решает реальную задачу — поиск и связывание знаний внутри организации.

Но для вертикального AI у горизонтальных платформ есть структурное ограничение: они не накапливают данные траекторий в конкретном операционном домене. Возьмём поток обращений в службу поддержки SaaS-продукта. Горизонтальная платформа знает, что в компании есть документация по продукту и тикеты. Она не знает, какие решения о приоритизации, эскалации или закрытии тикета в каком контексте приводили к удержанию клиента, а какие — к отказу от подписки в течение следующих 90 дней. Эта связь между решением и исходом не накапливается автоматически — она требует явной фиксации в момент действия и привязки к измеримому событию через N дней.

Исторический прецедент — Veeva против Salesforce в фармацевтике. Veeva не победила лучшей CRM-системой. Она победила потому, что накопила domain-specific онтологию и историю реальных взаимодействий именно для pharma sales: паттерны контакта с врачами, регуляторные ограничения, специфику цикла одобрения препарата. Salesforce не воспроизвёл это через универсальную CRM не потому, что не мог собрать данные, а потому, что схема и контекст у Veeva уже были откалиброваны. Sanjay Srivastava в Forbes-эссе «The Moat Is Moving» формулирует тот же тезис в более узком виде: institutional regulation, deep ecosystem integration и, главное, «closed-loop data — operational evidence linking decisions to outcomes» в 2026 году становятся не маркетинговой «defensibility», а единственными слоями, которые провайдер модели физически не может упаковать в свой SDK.

Этот механизм — data network effects, описываемый в a16z-эссе «The Economic Case for Generative AI and Foundation Models» как одна из ключевых форм defensibility в эпоху foundation моделей: ценность компаундируется, когда больше использования системы создаёт лучшую калибровку, а лучшая калибровка привлекает больше использования.

Разрыв в стоимости между вертикальным маховиком и горизонтальной платформой здесь не косметический, а структурный. Совокупная стоимость владения горизонтальной enterprise-платформой уровня Glean для крупной компании, по публичным выкладкам аналитиков рынка корпоративного поиска, измеряется сотнями тысяч долларов в год. Вертикальный AI-контур, собранный поверх open-source инфраструктуры памяти и собственной онтологии домена, на сопоставимом периметре существенно дешевле. Эта ценовая разница — не конкурентный риск для вертикального игрока, а его структурная защита:

горизонтальная платформа не может опуститься в эту цену, не обесценив свою же модель монетизации.

Что произойдёт, когда managed eval станет товаром?

Закономерный риск: OpenAI уже выпустил Evals API. Anthropic движется в ту же сторону. Если managed evaluation станет бесплатной инфраструктурой — не обесценивает ли это запись траекторий как таковую?

Нет — по той же причине, по которой Datadog существует рядом с бесплатным AWS CloudWatch. CloudWatch даёт сырые метрики; Datadog продаёт интерпретации в контексте конкретного стека и рабочего процесса. Gong существует рядом с Salesforce Einstein много лет по той же логике: Salesforce умеет анализировать структурированные данные CRM, но не воспроизводит специфику разговоров, которую Gong накопил в конкретных индустриях.

Почему это работает как порог, а не как движение

Накопление траекторий ведёт себя не линейно. До определённого объёма записей система ведёт себя как «лучший промпт» — её можно воспроизвести копией инструкции. После того, как система накопила заметный объём закрытых циклов в одном операционном контексте, «перенести знание» к конкуренту выписыванием промпта уже нельзя: ценность живёт в парах «решение → исход», в сезонных паттернах и в длинном хвосте исключений, которые нельзя выразить одной инструкцией. Конкуренту придётся накапливать это заново — в том же режиме реального времени, в том же домене, при тех же условиях.

Публичных бенчмарков, которые прямо транслируют размер trajectory корпуса в «проценты switching cost», пока нет — и это честнее признать. Но форма разрыва фиксируется количественно на смежных прокси. MemoryAgentBench (arXiv:2507.05257) показывает, что разрыв в качестве temporal reasoning между системами с накопленной историей и без неё измеряется десятками процентных пунктов на LongMemEval — и этот разрыв не закрывается ни ростом контекстного окна, ни сменой модели. Единственный фактор, который его закрывает, — накопление реальной истории в том же домене.

Без явного слоя записи траекторий — фиксации входа, решения и исхода — эти данные не накапливаются автоматически. Это инженерное решение, которое имеет смысл принимать на старте: вход (контекст), решение (что агент или человек сделал), исход (что произошло через N дней) — три поля в каждой записи. Ретроспективная разметка старых логов возможна, но контекст момента уже потерян, и качество таких данных принципиально ниже.

Что из этого следует

Слой	Что заменить	Сколько времени	Что нельзя воспроизвести
Модель	Любой альтернативный LLM	1-2 недели	-
Memory backend	Graphiti → MemO → Letta	2-4 недели	-
Схема данных	Написать новую онтологию домена	3-6 месяцев	Калибровку исключений и граничных случаев
Trajectory data	Нельзя	-	Тысячи реальных пар «решение → исход» (для иллюстрации)

Четыре следствия для вертикального AI-продукта.

Первое. Архитектурное решение: запись траекторий должна быть явным словом с первого дня, а не логом, который «потом разметим» (arXiv:2506.15841). Вход (контекст), решение (что агент или человек сделал), исход (что произошло через N дней) — три поля в каждой записи. Без этой структуры данные копятся, но не становятся обучающим сигналом.

Второе. Коммуникационное решение: клиент должен понимать, что именно компаундируется со временем. Не «мы поддерживаем AI-систему», а «за 6 месяцев система накопила N закрытых циклов в вашем домене, и это делает её точнее, чем любое новое решение с нуля». Gong продаёт ровно это сообщение: клиент платит не за записи звонков, а за то, что система с каждым месяцем точнее предсказывает исходы сделок, чем интуиция менеджера.

Третье. Стратегическое следствие: горизонт конкурентного преимущества зависит от того, как давно и насколько методично ведётся запись (arXiv:2507.05257). Провайдер, начавший фиксировать траектории год назад, имеет год структурного опережения. Этот разрыв не закрывается «лучшей моделью» — только временем работы в том же контексте. Оценка вертикального AI-продукта в этой логике строится не на технологическом стеке, а на глубине и длине накопленной истории.

Четвёртое. Конкурентное следствие: managed eval от OpenAI и Anthropic не уравнивает игровое поле — он смещает ценность дальше в сторону данных (Evals API). Когда инфраструктура записи доступна всем, выигрывает тот, кто начал записывать раньше и в правильном вертикальном контексте. Коммодитизация инструментов в зрелых рынках систематически усиливает преимущество тех, кто накопил содержание раньше — этот сдвиг ценности от

инструмента к данным в его работе уже прошли инфраструктурный мониторинг (CloudWatch против Datadog) и CRM (Salesforce против Gong); сейчас он повторяется на уровне AI-агентов.

Главное

- **Модель и слой памяти коммодитизируются** на горизонте 12–18 месяцев — это базовое условие рынка, на котором приходится строить продукт.
- **Схема данных (representation layer) держится годами.** Необходимый, но не достаточный уровень защиты.
- **Данные траекторий — единственный слой с физической невозпроизводимостью.** Закрытые циклы «решение → исход» нельзя купить, синтезировать или воссоздать ретроспективно (arXiv:2506.15841).
- **Поведение слоя нелинейно:** до определённого объёма записей знание переносится копией промпта; после — нет, и конкуренту приходится накапливать заново в реальном времени.
- **Рынок уже монетизирует этот слой:** Gong (~\$28 000–\$106 000/год на команду), Langfuse и Braintrust — отдельная позиция прайс-листа за хранение и анализ траекторий.
- **Managed eval не убивает преимущество** — он создаёт инфраструктуру записи. Ценность — в том, что именно записано и под какой домен откалибровано.

FAQ

Чем данные траекторий отличаются от обычных логов системы?

Лог фиксирует факт: агент ответил так-то в такое-то время. Запись траектории фиксирует замкнутый цикл: агент принял такое решение в таком контексте, и через N дней исход был таким. Разница принципиальная — лог описывает действие, запись траектории связывает действие с последствием. Эта связь и создаёт обучающий сигнал, который улучшает следующие решения в похожих ситуациях (arXiv:2506.15841).

Почему нельзя воссоздать данные траекторий задним числом — разве нельзя разметить старые логи?

Ретроспективная разметка работает для очевидных случаев, но не воспроизводит контекст момента решения. Временная метка, история предыдущих взаимодействий, состояние системы, параллельные события — всё это либо не сохранилось в логах, либо не было связано в структуру, пригодную для извлечения паттернов. MEM1 (arXiv:2506.15841) показывает, что внутренние представления модели, обученной на реальных траекториях, отличаются от представлений модели, обученной на синтетически восстановленных записях, — и этот разрыв не закрывается ростом объёма синтетики.

Горизонтальные платформы корпоративной памяти вроде Glean — разве они не решают ту же задачу?

Нет. Горизонтальная платформа знает, что в компании есть информация по домену, но не знает, какие решения в каком контексте приводили к каким исходам. Это разница между корпусом документов и историей действий, привязанных к последствиям. Горизонтальная платформа работает с первым; данные траекторий — это второе. Вертикальный AI-продукт с двенадцатью месяцами работы в домене и горизонтальная корпоративная память — разные категории, а не конкуренты.

Когда данные траекторий начинают создавать реальное конкурентное преимущество — с первого дня или нужно накопить порог?

Поведение нелинейно. До определённого объёма закрытых циклов в одном операционном домене знание системы можно перенести к конкуренту копией инструкции — switching cost практически нулевой. После — нет: ценность живёт в парах «решение → исход», в сезонных паттернах и в длинном хвосте исключений, которые нельзя выписать одной инструкцией. Конкурент в этой точке вынужден накапливать заново в режиме реального времени, в том же домене и при тех же условиях. Конкретный порог зависит от плотности решений и длины обратной связи в вертикали — публичных бенчмарков, переводящих размер корпуса траекторий в проценты switching cost, пока нет.

Как объяснить клиенту ценность данных траекторий, не раскрывая технических деталей?

Лучше всего работает аналогия с опытным сотрудником. Новый сотрудник может знать тот же регламент, что и опытный, но опытный помнит, что у определённого типа клиентов принято работать не по регламенту, что некоторые обращения требуют звонка до отправки ответа, что сезонный пик нужно прогнозировать заранее. Эти знания нельзя передать в инструкции — они накапливаются в работе. Данные траекторий — это способ, которым AI-система фиксирует и использует именно такой опыт.