

Charge metric: главное решение AI-компании, которое не пишут в прайс-листе

2026-05-20

Charge metric: главное решение AI-компании, которое не пишут в прайс-листе

Когда Salesforce выпустил первую версию Agentforce в конце 2024 года, ценообразование было простым: \$2 за разговор. Казалось логичным — агент работает, ты платишь за работу. Уже через несколько месяцев компания переделала схему, добавив гибридный вариант с кредитами, подписками и лицензиями на уровне отделов. Причина: разные покупатели внутри одной компании хотели принципиально разных гарантий предсказуемости.

Этот эпизод точно описывает, в каком переходном состоянии сейчас находится ценообразование AI-продуктов. Старая логика — «больше людей, больше мест, больше выручки» — строилась на предположении, что ценность ПО масштабируется с числом пользователей. AI-агент это предположение нарушает: один агент может выполнять объём работы, который раньше требовал десятка сотрудников. Если вы взяли за него \$X/место в месяц, вы проиграли сами себе.

\$X/seat/month — это модель ценообразования, при которой клиент платит фиксированную сумму за каждого человека (seat = лицензированного пользователя), получившего доступ к продукту. Она работает, пока добавление каждого нового пользователя приносит вендору пропорциональную выручку без пропорционального роста расходов. AI-агент ломает обе половины этого уравнения: один агент обслуживает работу многих пользователей, а его операционная стоимость растёт не с числом seats, а с объёмом выполненной работы. Этот сдвиг — продолжение более широкой темы роли важнее места: если сама работа перестаёт быть «местом», то и счёт нельзя выставлять за «места».

Откуда взялась модель «место в месяц» и почему она начинает давать сбои

В SaaS 2000-х годов лицензия за место была разумной прокси для ценности. Чем больше людей использует CRM — тем более ценным он становится для компании. Salesforce, HubSpot, Notion — все они стоят ровно столько, сколько число активных пользователей умножить на тариф. Модель работала, потому что программное обеспечение усиливало людей, но не замещало их.

Gross margin SaaS при этом держалась на уровне 80–90% — издержки на доставку продукта (хостинг, поддержка) были невысокими, а дополнительное место не требовало дополнительных переменных затрат. В этой конструкции место в месяц было почти идеальной единицей расчёта: предсказуемо для клиента, почти бесплатно для вендора.

AI-агенты работают по принципиально другой логике. Sierra, компания по автоматизации клиентского сервиса, заряжает за каждый разрешённый тикет, а не за число операторов, подключённых к платформе. Intercom Fin стоит \$0.99 за результат — то есть за каждый случай, когда бот закрыл обращение без передачи человеку. Логика очевидна: если ценность в том, что заявка обработана, а не в том, что оператор залогинился, почему единицей расчёта должен быть оператор?

Проблема не абстрактная. Компания, которая продаёт AI-агента для обработки входящих лидов, сталкивается с конкретным вопросом: сколько «мест» занимает агент? Один? Эквивалент пяти менеджеров по продажам, которых он заменяет? Ответа у per-seat логики нет.

Более того, клиент, который покупает AI-агента по per-seat модели, неизбежно начнёт давить на цену, считая «справедливой» ценой долю стоимости замещаемых сотрудников. Это ставит вендора в позицию, где потолок выручки определён числом людей, которых агент заменяет, — а это потолок из мира кадрового аутсорсинга, а не SaaS.

Три слоя ценообразования, которые победили

Bessemer Venture Partners в своём AI Pricing Playbook, опубликованном в начале 2026 года, выделяют три принципиально разных способа посчитать ценность AI-продукта:

| Модель | Что считается | Пример | Маржинальность |
|--|----------------------------|--|----------------|
| Потребление (токены, API-вызовы) | Вычислительные ресурсы | Cognition Devin — \$2.25 за единицу вычислений (асу) | 60–70% |
| Процессы (задачи, операции) | Завершённые единицы работы | Salesforce Agentforce — кредиты | 50–65% |
| Результат (исход, решение) | Бизнес-итог | Sierra — \$X/resolved ticket | 40–60% |

Как подчёркивает Bessemer: по мере движения от потребления к результату вы берёте на себя больше стоимостного риска в обмен на более точное выравнивание ценности. Платёж за результат идеально объясняет клиенту, почему он платит, — но требует от вендора понимать свои операционные расходы с точностью до единицы.

По данным аналитиков Hire Fraction, 92% AI-компаний уже применяют смешанные схемы — базовая подписка плюс потребление или результат. Чистый per-seat остаётся только там, где агент действительно усиливает существующего сотрудника, а не замещает его. Таких ситуаций становится меньше.

Как это выглядит у конкретных компаний

Harvey: дорогие места с принудительным расширением

Harvey — AI-ассистент для юридических фирм — продаёт по \$1 200–2 500 за место в месяц. По меркам SaaS это дорого. По меркам юридического рынка — не очень: один ассоциат в BigLaw обходится фирме в \$200–400k в год.

Согласно данным Sacra, Harvey вырос от нуля до \$195 млн ARR за 36 месяцев при 290% роста год к году. Ключ не в том, что место стоит дорого, а в том, как оно расширяется: компания начинает с одной практики (например, M&A), затем добавляет Workflows, затем переходит на другие подразделения. Минимальный порог входа — 20 мест и 12-месячный контракт, то есть минимальный годовой контракт от \$288k.

Здесь per-seat работает, потому что место — это место юриста, который всё равно нужен. Агент усиливает, а не заменяет. Но выручка растёт не через новые места, а через expansion внутри клиента.

Sierra: чистый результат

Компания Bret Taylor и Clay Bavor пошла по другому пути. Sierra продаёт агентов для клиентского сервиса, и её ценообразование — это платёж за результат без фиксированной подписки. Минимальный годовой контракт около \$150k, но структура — исключительно за разрешённые обращения.

В блоге компании Sierra объясняет логику прямо: «Мы берём деньги только тогда, когда клиент получает реальную ценность». Это жёсткое утверждение, потому что оно привязывает выручку вендора к измеримому качеству работы агента. Если агент начинает хуже разрешать тикеты — падает выручка.

Glean: ладдер от мест к агентной нагрузке

Glean, платформа корпоративного поиска с AI, использует гибридную модель: базовая подписка \$45–50/место в месяц, дополнительный AI-модуль +\$15/место, а поверх этого — consumption-based SKU для агентных рабочих нагрузок. По данным Futurum, компания удвоила ARR с \$100M до \$200M за девять месяцев.

Это иллюстрация «безопасного» перехода: не нужно отказываться от per-seat полностью. Достаточно добавить новые SKU поверх существующей базы, которые монетизируют агентное поведение отдельно.

Что меняется в логике сделки

За сменой единицы расчёта стоит более глубокий сдвиг. Когда клиент платит за место, договор звучит так: «вы купили доступ к инструменту — используйте его хорошо». Когда клиент платит за результат, договор другой: «мы берём деньги только тогда, когда работа сделана».

Это перекладывает ответственность. Вендор на per-seat не несёт ответственности за то, что менеджер открыл систему пять раз за месяц. Вендор на

per-outcome несёт ответственность за каждый нерешённый тикет — в прямом финансовом смысле.

Foundation Capital в своём эссе о service-as-software формулировали этот принцип как SaaS²: AI-компании могут захватить в 10 раз больше ценности, чем традиционный SaaS, именно потому что они выполняют работу, а не просто дают инструмент для её выполнения. Но это работает только если ценообразование отражает работу, а не доступ к инструменту.

Обратная сторона этого договора — вендор начинает нести ответственность за качество работы. При per-seat клиент несёт ответственность за то, как использует продукт. При per-outcome ответственность частично переходит к поставщику: если агент работает плохо, выручка вендора снижается автоматически. Это создаёт принципиально другую мотивацию к качеству продукта.

Как это читается для трёх аудиторий

Для основателя AI-продукта. Вопрос «за что мы берём деньги» — это не строка в прайс-листе, а стратегическое заявление о том, где вы видите ценность. Bessemer называют выбор «charge metric» самым важным решением AI-компании. Если ваш агент замещает сотрудников — per-seat будет работать против вас: клиент не купит пять мест для одного агента. Найдите результат, который клиент уже измеряет, и зарядите за него.

Для операционного директора компании, покупающей AI. Чистый per-seat с AI-поставщиком — это красный флаг. Либо инструмент просто усиливает существующих сотрудников (тогда per-seat оправдан), либо он что-то делает сам — и тогда вы должны понимать, за какую именно работу платите. Требуйте прозрачной метрики результата в контракте.

Для product manager'a. Если ваш продукт переходит от ассистентского режима (AI помогает человеку) к автономному (AI делает сам), это требует ревизии юнит-экономики. Gross margin у outcome-based AI составляет 40–60% против 80–90% у классического SaaS — значит, стоимость операционных расходов должна быть под контролем уже сейчас.

Сигналы, на которые стоит смотреть

Рынок ещё не устоялся. Salesforce переделала Agentforce дважды за год. Cognition Devin упал в цене на порядок между первым и вторым запуском. Это не признаки слабости — это признаки того, что правильная единица расчёта для разных типов агентных продуктов только нащупывается.

Две вещи, которые стоит отслеживать: первая — как крупные enterprise-покупатели начнут требовать стандартизации метрик результата (сейчас «resolved ticket» у каждого поставщика значит что-то своё). Вторая — появятся ли revenue-share схемы в вертикальных сегментах, где AI-агент напрямую участвует в генерации выручки клиента. Harvey уже работает с такими договорённостями в юридическом секторе.

Что это значит для российского контекста

В РФ дискуссия про outcome-based pricing пока редкость — большинство B2B AI-проектов продаются либо как setup-проект с фиксированной сметой, либо как ежемесячный retainer. Это объяснимо: outcome-based требует трёх вещей, которых на рынке пока мало: согласованных метрик результата, доверия между поставщиком и клиентом и инфраструктуры для подсчёта тех самых результатов.

Но направление движения то же. Заказчик, который заплатил 800 тысяч рублей за «внедрение AI-ассистента менеджеров», через три месяца спрашивает не «сколько раз бот залогинился», а «сколько лидов он квалифицировал и сколько встреч сгенерировал». То есть в голове у клиента метрика уже outcome-based, даже если в договоре стоит фиксированная сумма.

Это создаёт окно для тех, кто готов первыми сшить контракт под результат. Не обязательно идти в чистый per-outcome — для большинства русских клиентов это слишком радикально на старте. Гибрид «base retainer + bonus за достигнутый KPI» куда продаваемее, потому что сохраняет предсказуемость для финансиста и при этом привязывает выручку поставщика к качеству работы.

Здесь же возникает структурный вопрос: setup-фаза, в которой строится сам контур, уже плохо масштабируется и постепенно перестаёт быть точкой захвата маржи. Если ценность переместится в operational layer — туда же должна переместиться и единица расчёта. Retainer привязанный к доступу — это всё ещё SaaS-логика; retainer, привязанный к delivered work, — уже outcome-based.

Почему рынок ещё не перешёл на outcome-based: чего не хватает?

Два технических блока удерживают рынок на per-seat дольше, чем стоило бы.

Первое — атрибуция результата. Когда тикет «разрешён» — это решил агент, или человек, дочистивший его за агентом? Когда лид «квалифицирован» — это работа бота или менеджера, который через час перезвонил? Без чёткой attribution-системы любой outcome-based контракт превращается в спор о том, кто что сделал. Sierra тратит непропорционально много инженерных усилий именно на это: их платформа умеет показывать клиенту, что конкретно сделал агент, а что — оператор.

Второе — наблюдаемость экономики на стороне вендора. Per-seat предсказуем по определению: подписали 200 мест — знаешь выручку на год вперёд. Outcome-based требует, чтобы вендор в реальном времени видел свою себестоимость на единицу результата: сколько inference-вызовов, сколько эскалаций к человеку, сколько повторных обращений. Большинство AI-стартапов эту инфраструктуру не строят на старте — и попадают в ловушку, когда выручка есть, а маржинальность отрицательная.

Эти два блока — не вспомогательные. Они и есть стек, который должен быть готов, прежде чем компания может перейти от per-seat к outcome-based. Без них переход даёт обратный эффект: выручка падает, риски растут.

Главное

- Модель «место в месяц» строилась на том, что ценность ПО масштабируется с числом пользователей. AI-агенты это допущение нарушают.
- Bessemer выделяют три уровня ценообразования: потребление → процесс → результат. По мере роста «выравнивания ценности» растёт и стоимостной риск для вендора.
- 92% AI-компаний уже используют гибридные схемы; чистый per-seat остаётся только там, где агент усиливает, а не замещает.
- Harvey (\$195M ARR, +290% в год) зарядил дорого за место, но растёт через expansion внутри клиента. Sierra и Intercom зарядили за результат и убрали место из уравнения полностью.
- Смена единицы расчёта — это смена договора с клиентом: от «доступа к инструменту» к «оплате за работу».

FAQ

Чем «ценообразование за результат» отличается от «ценообразования за использование»?

Использование измеряет то, что система потратила (токены, вызовы API, вычислительные ресурсы). Результат измеряет то, что клиент получил (тикет закрыт, лид квалифицирован, документ подписан). Использование хорошо предсказуемо для вендора, но плохо объясняет ценность клиенту. Результат лучше выравнивает интересы, но требует чёткого определения того, что именно считается результатом.

Когда per-seat всё ещё оправдан для AI-продукта?

Когда агент действительно усиливает существующего сотрудника, а не выполняет работу вместо него. Harvey — корректный пример: юрист всё равно нужен, агент делает его быстрее. Если же продукт полностью заменяет функцию — per-seat работает против вендора: клиент видит агента как замену N мест и оценивает «справедливую» цену соответственно.

Как клиенту защититься от непредсказуемых счетов при outcome-based ценообразовании?

Ключевые параметры для переговоров: чёткое определение «результата» (кто считает и как), минимальный базовый платёж (предсказуемый floor), максимальный cap на расходы за период. Intercom Fin стоит \$0.99/результат, но есть минимум \$49.50/мес — это и есть floor + per-unit структура.

Насколько снижается маржинальность при переходе на outcome-based?

По данным Bessemer, классический SaaS работает на gross margin 80–90%. AI-продукты с outcome-based ценообразованием — 40–60%. Разница объясняет-

ся тем, что операционные расходы (inference, мониторинг, fallback на человека) теперь входят в стоимость «результата». Это не плохо — это другая бизнес-модель с другими мультипликаторами.

Что означает «charge metric — это стратегическое заявление»?

Bessemer формулируют так: выбор метрики расчёта определяет не только строку в инвойсе, но и то, какую проблему вы публично обязуетесь решать. Если вы берёте за место — вы говорите «мы продаём доступ». Если за результат — «мы продаём работу». Последнее сложнее продать на входе, но создаёт более прочные долгосрочные отношения с клиентом.

Можно ли начать с per-seat и потом перейти на outcome-based?

Можно, и так делают почти все, кто проходит этот переход осознанно. Glean — каноничный пример: per-seat остаётся базой, поверх неё постепенно появляются consumption-SKU за агентные нагрузки. Через 12–18 месяцев outcome-часть может вырасти до половины контракта. Резкий переход «было per-seat, стало per-outcome» обычно ломает существующие сделки — лучше двигаться слоями.