

# УПРАВЛЯЕМЫЕ ИИ-АГЕНТСКИЕ ПЛАТФОРМЫ КАК ИНФРАСТРУКТУРА ИЗМЕРИМОГО ЭКОНОМИЧЕСКОГО ЭФФЕКТА

**архитектура, доказательная база, ROI-модель и инвестиционная логика на  
примере Century**

Е. С. ПИСКУН, А. А. АЗИЗОВ, Е. В. КРЯЧЕВ

*Научно-аналитическая работа*

Версия с расширенной экономической частью, кейсами и расчетами ROI

## **Аннотация**

Работа исследует экономический эффект ИИ-агентских платформ как результат управляемого изменения бизнес-процесса, а не как следствие применения языковой модели в изоляции. Центральная научная задача состоит в построении атрибуционно корректной модели оценки: опубликованные внешние эффекты не присваиваются платформе Century напрямую, а используются как калибровочные бенчмарки, переносимые через коэффициенты готовности процесса, сходства кейса, монетизации и governance. Платформа Century рассматривается как перспективная case-model enterprise agentic platform, объединяющая ассистентов, документы и знания, workflow, ролевой доступ, аудит, метрики и варианты on-prem, air-gapped и hybrid-развертывания. Экономический блок дополнен расчетами на основе реальных публичных кейсов: Klarna, С.Н. Robinson, Accenture/Celonis, UPS ORION, Microsoft contact centers, а также авторской риск-финансовой модели IMA. В отсутствие закрытых логов, чеков внедрения и фактических пилотных метрик Century фактический ROI платформы не декларируется; вместо этого рассчитаны нормализованные эффекты, пороги окупаемости, DCF/NPV, сценарии внедрения и инвестиционные критерии, применимые для будущего пилота.

Ключевые слова: ИИ-агенты; agentic AI; Century; LLM; RAG; business process management; process mining; workflow; Human-in-the-Loop; ROI; NPV; ALE; enterprise AI; инвестиционная оценка.

## **1. Постановка проблемы и научная ценность работы**

Главная проблема корпоративного применения генеративного ИИ состоит не в отсутствии моделей, а в отсутствии измеримого и воспроизводимого экономического результата. Расширение AI-use сопровождается трудностью перехода от пилотов к масштабному эффекту: отчет McKinsey State of AI 2025 описывает рост применения AI и agentic AI, но одновременно указывает на незавершенность перехода от пилотов к scaled impact в большинстве организаций [10]. Макрорыночный фон усиливает давление на капитал: Gartner прогнозирует мировые расходы на AI почти в 1,5 трлн долл. США в 2025 г., а IDC оценивает AI-spending в 632 млрд долл. США к 2028 г. [7; 8]. Однако эти прогнозы не доказывают доходность конкретной платформы, а лишь подтверждают стратегический масштаб рынка.

Научная ценность работы заключается в переходе от обзорного сопоставления AI-кейсов к модели экономической атрибуции. В большинстве публичных материалов об agentic AI преобладают технические метрики, а не денежные результаты. Систематический обзор 84 работ 2023-2025 гг. показывает дисбаланс: технические метрики присутствуют в 83% оценок, тогда как экономические оценки встречаются только в 30%, а одновременно технические и human-centered измерения представлены лишь в 15% исследований [31]. Следовательно, необходима модель, которая связывает технический результат с процессными KPI, денежным эффектом, риском и ограничениями переноса внешних бенчмарков.

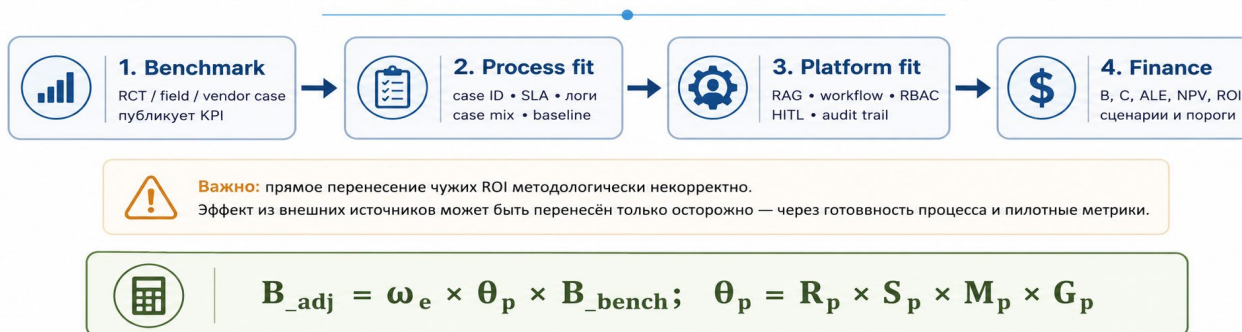
Предлагаемая рамка строится вокруг принципа attribution-safe ROI. Внешний кейс используется не как доказательство эффективности Century, а как источник наблюдаемого эффекта по аналогичному типу процесса. Перенос допускается только после оценки готовности процесса, сходства с исходным кейсом, монетизируемости эффекта и зрелости governance. Такой подход предотвращает методологическую ошибку, при которой экономический результат Klarna, UPS или Accenture автоматически приписывается иной платформе и иной организации.

$$B_p^{adj} = \omega_e \cdot \theta_p \cdot B_p^{bench}, \quad \theta_p = R_p \cdot S_p \cdot M_p \cdot G_p$$

(1) модель осторожного переноса внешнего бенчмарка в процесс p

В формуле (1)  $B_{adj,p}$  означает скорректированную выгоду процесса p;  $B_{bench,p}$  - опубликованный бенчмарк;  $\omega_e$  - коэффициент доказательности источника;  $\theta_p$  - коэффициент переносимости;  $R_p$  - готовность процесса по данным, логам и регламентам;  $S_p$  - сходство с исходным кейсом;  $M_p$  - доля эффекта, реально превращаемая в денежный поток;  $G_p$  - зрелость governance, включая RBAC, аудит и Human-in-the-Loop. Для Century такая модель особенно важна из-за отсутствия закрытых пилотных логов: количественные эффекты могут быть указаны только как расчетные сценарии, а не как фактические достижения платформы.

### Научная рамка: перенос эффекта без присвоения чужих результатов



Ценность модели состоит в осторожной экономической атрибуции: Century оценивается через готовность процесса и пилотные данные, а не через прямое перенесение чужих ROI.

Рисунок 1. Научная рамка attribution-safe ROI: перенос внешнего эффекта только через готовность процесса и пилотные метрики.

## 2. Century как case-model управляемой агентской платформы

Century целесообразно анализировать не как чат-бот и не как отдельный LLM-интерфейс, а как потенциальную платформу управляемой процессной агентности. В публичном описании Century указано, что платформа объединяет ассистентов, работу с документами и знаниями, workflow, контроль доступа, аудит и эксплуатационные метрики в одном контуре; дополнительно заявлена пригодность для on-prem, air-gapped и hybrid-развертывания [4]. Отдельное описание Stacklevel Group подчеркивает атрибуцию источников, учет прав доступа, аудиторские журналы, этапы оценки до релиза и в эксплуатации, а также локальный, изолированный и гибридный deployment [5].

Экономическое значение такой архитектуры связано с тем, что в регулируемых и документно-интенсивных организациях сам факт генерации ответа не является достаточным активом. Активом становится управляемый сценарий: документ загружен и очищен, источник проверен, права доступа применены, фрагменты базы знаний процитированы, промежуточные шаги зафиксированы, критические действия переданы в Human-in-the-Loop, итоговый результат связан с KPI процесса. Именно эта логика отличает агентскую платформу от обычного корпоративного ассистента.

Сценарии Century, описанные на публичных страницах платформы, релевантны для политик и регламентов, документов и договоров, комплаенса и операционных процессов, где требуется проверяемость, учет прав доступа и управляемый маршрут исполнения [6]. В таких сценариях источниками экономического эффекта выступают не только скорость ответа, но и уменьшение числа

ошибок, снижение повторных обращений, сокращение времени согласования, снижение стоимости контроля и уменьшение риска некорректного доступа к данным.

## Century как case-model управляемой агентской платформы



Рисунок 2. Техническая логика Century как case-model: знания, workflow, агенты и governance в едином контуре.

### 3. От одиночного ИИ к агентской платформе: сравнительная экономическая логика

Для экономического анализа принципиально различаются три режима цифровизации процесса. Первый режим - процесс без ИИ, в котором классификация, поиск информации, заполнение документов, согласования и ответы клиенту выполняются человеком. Второй режим - обычный AI-assist, когда LLM ускоряет отдельный шаг, но маршрут кейса, статусы, действия в системах и контроль остаются в основном ручными. Третий режим - агентская платформа, где AI связан с корпоративными источниками, workflow, правами доступа, инструментами, журналом действий и KPI.

Режим	Операционная характеристика	Что измеряется	Экономическая атрибуция
Без ИИ-платформы	Ручной процесс; знания находятся в документах и памяти сотрудников	Время кейса, очередь, ошибки ручного переноса, SLA-просрочки	Низкая: эффект изменений часто неотделим от состава кейсов
Обычный LLM / AI-assist	Модель ускоряет текст, поиск, резюмирование или подсказку оператору	Локальная производительность сотрудника; например +14% issues/hour в field experiment [11]	Средняя: полезно для отдельного шага, но workflow и контроль не становятся автоматическими
Агентская платформа	Данные, RAG, workflow, tool access, RBAC, HITL и audit trail работают как единый контур	Время кейса, повторные обращения, throughput, SLA, риск, cost per case, NPV	Высокая при наличии baseline, логов и control-группы; именно здесь возникает платформенный ROI

Таблица 1. Отличие обычного AI-assist от агентской платформы для целей ROI-оценки.

Ключевой вывод состоит в том, что обычный LLM может показывать высокий эффект по времени выполнения отдельной задачи, но агентская платформа имеет более широкий экономический профиль: она меняет маршрут процесса, снижает повторную работу, создает audit trail и позволяет управлять правами, рисками и стоимостью каждого сценария. В финансовой модели это означает переход от локального показателя productivity uplift к совокупному процессному эффекту.

$$E_{net,p} = B_{time,p} + B_{quality,p} + B_{SLA,p} + B_{risk,p} + B_{rev,p} - C_{int,p} - C_{plat,p} - C_{inf,p} - C_{HITL,p} - C_{chg,p}$$

(2) сетевой процессный эффект агентского сценария  $p$

В формуле (2)  $B_{time}$  отражает высвобождение времени и рост пропускной способности;  $B_{quality}$  - снижение ошибок, дефектов и повторных обращений;  $B_{SLA}$  - экономию на просрочках и очередях;  $B_{risk}$  - предотвращенные потери и снижение ALE;  $B_{rev}$  - дополнительный доход от ускорения или роста конверсии. Блок затрат включает интеграцию, платформу, инференс, Human-in-the-Loop и организационные изменения.

#### 4. Расширенная доказательная база: кейсы с реальными цифрами

Доказательная база должна разделяться по уровню надежности. Рандомизированные эксперименты и field studies полезны для оценки причинного эффекта в ограниченном процессе; vendor/customer cases дают ценные операционные цифры и денежные эффекты, но требуют осторожности из-за отсутствия raw logs и контрольных групп. В приложенном исследовательском файле такие источники специально разделены по надежности: High, Medium/High, Medium/Low и Low/Medium [3].

В клиентской поддержке наиболее важен field experiment Brynjolfsson, Li, Raymond: выборка 5179 операторов показала рост производительности на 14% в среднем и до 34% у новичков и менее квалифицированных сотрудников [11]. Это бенчмарк для обычного AI-assist, потому что ИИ помогает сотруднику решать обращения, но не обязательно заменяет весь workflow. Эксперимент Nou и Zhang по knowledge-work зафиксировал сокращение времени выполнения письменных профессиональных задач на 40% и рост качества на 18% [12]. GitHub Copilot ускорил выполнение целевой задачи разработки на 55,8% [13]. Microsoft Security Copilot показал для опытных аналитиков 22% ускорения и 7% роста accuracy, а также был оценен в RCT-дизайне [14].

Корпоративные кейсы показывают уже операционный масштаб и денежные эффекты. Klarna сообщила, что AI-assistant в первый месяц обработал 2,3 млн диалогов, около двух третей customer-service chats, выполнял эквивалент работы 700 full-time agents, снизил повторные обращения на 25%, уменьшил время решения с 11 минут до менее чем 2 минут и, по оценке компании, должен был дать 40 млн долл. США улучшения прибыли в 2024 г. [16]. С.H. Robinson сообщила об автоматизации более 10 000 email transactions в день, сокращении обработки emailed load tender с периода до 4 часов до 90 секунд, а также о 5500 shipment orders в день за 90 секунд, 2600 quotes в день за 32 секунды и 3000 appointments в день менее чем за 60 секунд [17].

В P2P/procurement Accenture и Celonis представили кейс, где процессная прозрачность и execution-приложение дали 35 млн долл. США annualized working-capital benefits, 50% improvement in request-to-order cycle time, 30% reduction in invoice approval time, а в отдельном market unit цикл согласования был сокращен с 60 до 15 часов [18]. В логистике UPS ORION является важным примером процессного AI/operations research с реальной экономикой: по данным INFORMS, проект стоил около 250 млн долл. США, уже к декабрю 2015 г. сэкономил более 320 млн долл., а при полном внедрении ожидалось 300-400 млн долл. ежегодной экономии, 100 млн миль и 10 млн галлонов топлива в год [21].

Отдельно следует выделить авторскую работу по IMA. В ней на SecBench-25 показано снижение ASR с 16,0% до верхней границы менее 0,3%, что соответствует снижению ALE на 98,1%; средняя задержка в атакующих сценариях сокращалась на 52,7% за счет раннего отсека вредоносных запросов [2]. В экономической части приведен сценарий со 1000 активных пользователей, 5 млн запросов в год, InitialEx 150 тыс. долл., OpEx 180 тыс. долл. в год, предотвращенными потерями 244 тыс. долл. в год и NPV 80,7 тыс. долл. при  $r = 12\%$  [2]. Этот кейс важен для Century не как прямой компонент платформы, а как доказательство необходимости включать risk-adjusted savings в ROI агентных систем.

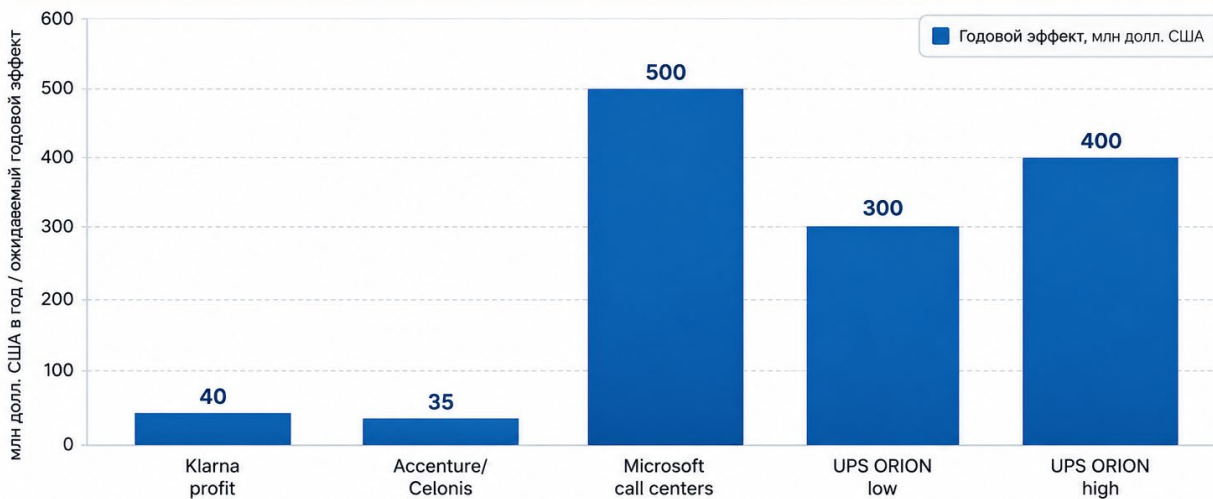
Кейс	Процесс	Опубликованный операционный эффект	Денежный эффект / ROI-значение
Klarna	Customer support	2,3 млн диалогов за 1 месяц; 11 мин -> <2 мин; -25% повторных обращений; эквивалент 700 FTE	Оценка компании: +40 млн долл. прибыли в 2024 г. [16]
C.H. Robinson	Логистика, email transactions	>10 тыс. транзакций в день; load tender до 4 ч -> 90 сек; 5500 orders/day	Денежный эффект не раскрыт; публикуются время и масштаб [17]
Accenture / Celonis	P2P, procurement, invoices	60 ч -> 15 ч в market unit; 50% request-to-order cycle time; 30% invoice approval time	35 млн долл. annualized working-capital benefits [18]
Microsoft	Contact centers	AI-сбережения в call centers, reported by Bloomberg/Reuters	>500 млн долл. экономии за прошлый год, по сообщению Reuters/Bloomberg [22]
UPS ORION	Route optimization	100 млн миль и 10 млн галлонов топлива в год; 6-8 miles per route/day	Стоимость проекта 250 млн долл.; ожидаемая экономия 300-400 млн долл./год [21]
IMA	LLM-agent security	ASR 16,0% -> <0,3%; ALE -98,1%; latency under attack - 52,7%	NPV 80,7 тыс. долл.; PI 1,54; DPP около 2,9 года [2]

Таблица 2. Расширенные кейсы с денежными и процессными показателями.

## Опубликованные денежные эффекты: только публично заявленные величины



Данные основаны исключительно на публичных источниках компаний. Не являются результатами внедрения платформы Century.



Что показывает график: порядок величины годового денежного эффекта по публичным кейсам.



Источники кейсов: публичные отчеты и пресс-релизы компаний (2023–2025 гг.).



Важно: перенос эффекта возможен только через модель атрибуции и оценку готовности процесса.

Рисунок 3. Публично заявленные денежные эффекты; показатели не являются результатами Century и используются как бенчмарки.

## 5. Нормализованное сравнение процесса: без AI, обычный AI и агентская платформа

Для сопоставления режимов внедрения необходима нормализация. В клиентской поддержке исходной точкой может служить опубликованное Klarna значение 11 минут на решение обращения до AI-assistant [16]. Для обычного AI-assist применяется field benchmark Brynjolfsson, Li, Raymond: +14% производительности в среднем и +34% у новичков [11]. Для агентного контура используется опубликованный Klarna уровень менее 2 минут; в расчетах принят консервативный потолок 2 минуты [16].

Режим	Минут/кейс	Часы на 100 тыс. кейсов	Сэкономлено часов	Индекс трудоемкости	Источник / расчет
Без AI	11.00	18 333	0	100%	Ручной baseline по Klarna: 11 минут [16]
LLM-assist, средний эффект	9.65	16 082	2 251	87,7%	Расчет из +14% issues/hour [11]
LLM-assist, эффект у новичков	8.21	13 682	4 652	74,6%	Расчет из +34% productivity [11]
Agentic workflow, уровень Klarna	2.00	3 333	15 000	18,2%	2 минуты и -25% repeat inquiries [16]

Таблица 3. Нормализованная трудоемкость customer support на 100 тыс. обращений.

## Нормализованное сравнение клиентской поддержки

чем ниже показатель, тем выше эффективность

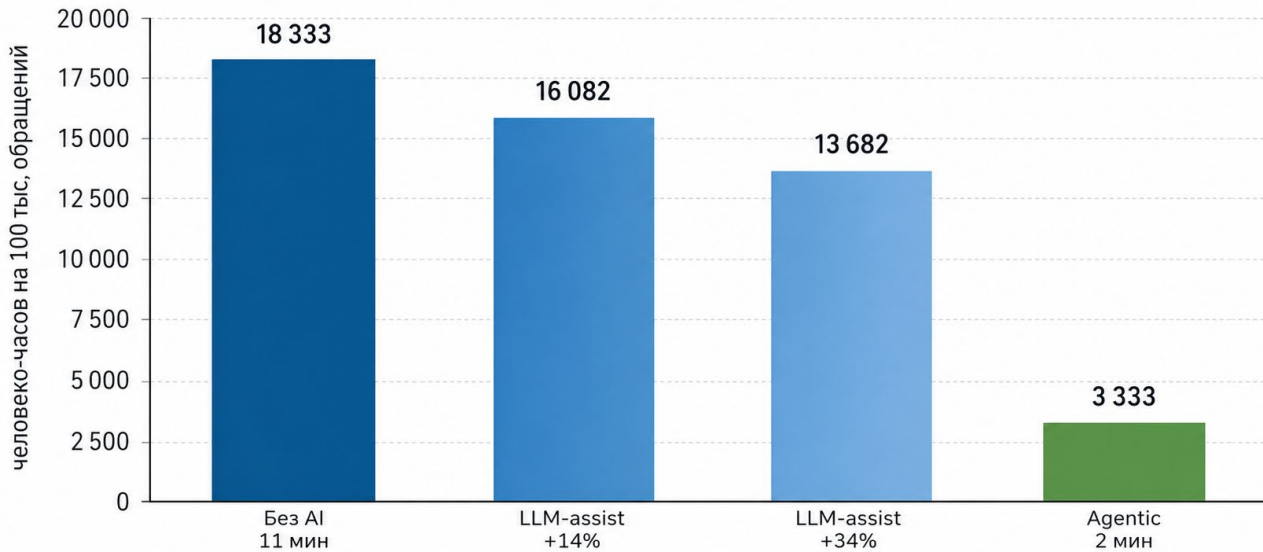


Рисунок 3. Нормализованное сравнение клиентской поддержки: по мере роста зрелости AI-поддержки трудозатраты снижаются.

Рисунок 4. Сравнение трудоемкости на 100 тыс. обращений: ручной процесс, LLM-assist и agentic workflow.

Монетизация нормализованного эффекта зависит от стоимости часа операционного труда и от того, превращается ли высвобождение времени в денежный поток. При 100 тыс. обращений агентный сценарий уровня Klarna высвобождает около 15 тыс. часов относительно ручного baseline. При условной полной стоимости часа 15, 30 и 50 долл. денежный эквивалент времени составляет 225 тыс., 450 тыс. и 750 тыс. долл. соответственно на каждые 100 тыс. обращений. Для обычного LLM-assist со средним эффектом +14% аналогичные значения составляют около 33,8 тыс., 67,5 тыс. и 112,6 тыс. долл. Эти величины являются расчетом на основе опубликованных time/productivity metrics, а не опубликованной прибылью Century.

Режим	Экономлено часов	Эквивалент при \$15/ч	Эквивалент при \$30/ч	Эквивалент при \$50/ч
LLM-assist +14%	2 251	33 772	67 544	112 573
LLM-assist +34%	4 652	69 776	139 552	232 587
Agentic workflow 2 мин	15 000	225 000	450 000	750 000

Таблица 4. Денежная чувствительность эффекта времени на 100 тыс. обращений.

Стоимость часа в таблице 4 является сценарным параметром. Реальные ставки оплаты, структура аутсорсинга, налоги, overhead и возможность фактического высвобождения FTE должны фиксироваться в пилоте конкретной организации.

В логистическом процессе С.Н. Robinson показывает еще более выраженную разницу между обычной цифровизацией и агентной автоматизацией. До нового решения emailed load tender мог занимать до 4 часов elapsed time; после внедрения - 90 секунд [17]. Если рассматривать только прямую ручную операцию ввода, компания указывает 7 минут на один load; сокращение до 1,5 минуты означает снижение прямой трудоемкости на 78,6%. При 5500 shipment orders в день это дает около 504 часов эквивалентной ручной работы в день, или около 126 тыс. часов в год при 250 рабочих днях. При ставке 30 долл./ч чувствительность равна примерно 3,78 млн долл. в год. Это расчетная монетизация опубликованных time/volume metrics, а не раскрытая прибыль С.Н. Robinson.

## 6. Экономическая модель: ROI, NPV, DPP и пороги окупаемости

Экономическая часть агентской платформы должна объединять операционный эффект, риск и стоимость владения. Базовая формула ROI не достаточна без учета того, что часть эффекта является

денежным потоком, часть - высвобожденным временем, часть - снижением риска, а часть - ускорением оборотного капитала. Поэтому расчет строится в двух слоях: сначала определяется сетевой эффект процесса  $E_{net,p}$ , затем он включается в денежный поток проекта с учетом первоначальных инвестиций, ежегодных затрат и ставки дисконтирования.

$$ROI_p = \frac{\sum B_p - \sum C_p}{\sum C_p}, \quad NPV = -I_0 + \sum_{t=1}^H \frac{CF_t}{(1+r)^t}$$

(3) базовые показатели доходности и дисконтированного эффекта

$$C_{ROI}^{max} = \frac{B}{1 + ROI^*}, \quad DPP = \min\left\{t: \sum_{i=1}^t \frac{CF_i}{(1+r)^i} \geq I_0\right\}$$

(4) пороговая стоимость платформы и дисконтированный срок окупаемости

Формула (4) особенно важна для Century в отсутствие коммерческих данных. Если фактическая стоимость пилота, лицензий, инфраструктуры и сопровождения не раскрыта, корректнее рассчитывать не мнимый ROI, а максимально допустимую all-in стоимость при заданной целевой доходности. Такой подход позволяет инвестору и клиенту оценить, в каких масштабах процесса внедрение становится экономически рациональным.

Бенчмарк	Годовой эффект, млн долл.	Стах при ROI 100%	Стах при ROI 200%	Стах при ROI 300%
Klarna-style support	40	20.0	13.3	10.0
Accenture/Celonis P2P	35	17.5	11.7	8.8
Microsoft call centers	500	250.0	166.7	125.0
UPS ORION low case	300	150.0	100.0	75.0

Таблица 5. Пороговая all-in стоимость внедрения при заданном ROI на основе опубликованных денежных эффектов.

### Пороговая экономика: сколько может стоить платформа при заданном ROI

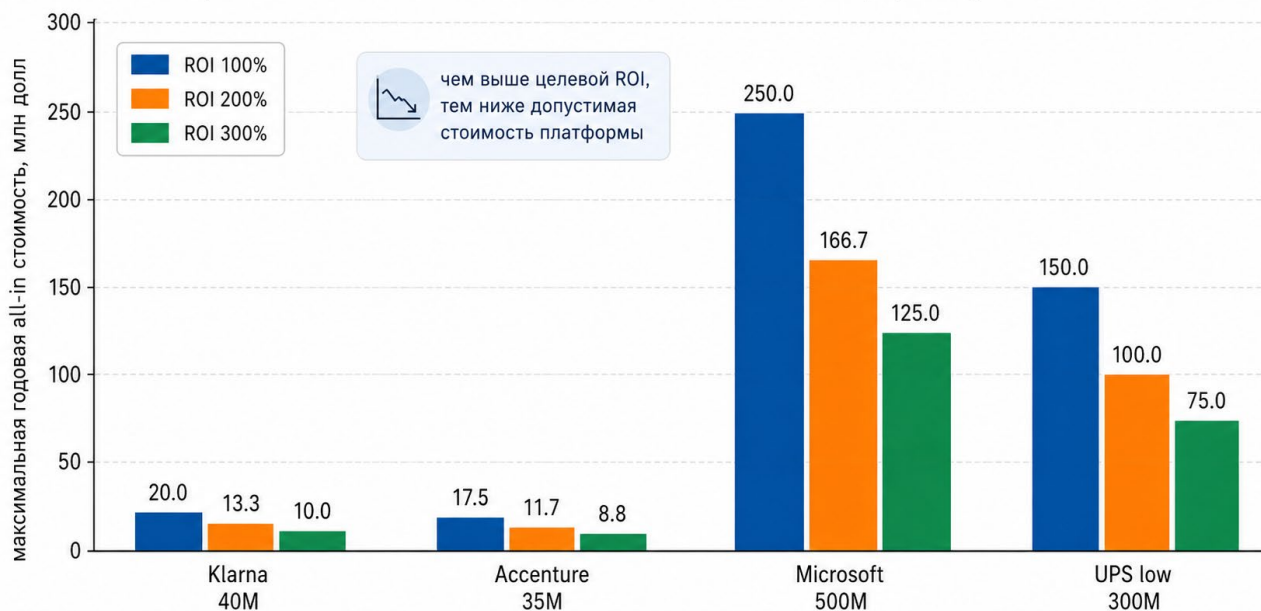


Рисунок 3. Пороговая экономика: допустимая all-in стоимость платформы при разных целевых ROI.

Рисунок 5. Максимальная годовая стоимость платформенного контура при целевых ROI 100%, 200% и 300%.

Таблица 5 показывает, что при benefit 40 млн долл. по аналогии с Klarna-style support all-in стоимость платформы должна быть не выше 20 млн долл. в год для ROI 100%, не выше 13,3 млн долл. для ROI 200% и не выше 10 млн долл. для ROI 300%. Для процесса Accenture/Celonis с 35 млн долл. annualized working-capital benefits соответствующие пороги составляют 17,5, 11,7 и 8,8 млн долл. Это не оценка цены Century, а управленческий фильтр: если ожидаемый масштаб процесса меньше, платформа

должна иметь меньший all-in cost или дополнительный эффект за счет портфельного переиспользования сервисов.

UPS ORION позволяет показать полноценный ROI на реальном крупном процессе, поскольку опубликованы и стоимость проекта, и ожидаемая экономия. При стоимости разработки и внедрения около 250 млн долл. и ожидаемой ежегодной экономии 300-400 млн долл. простой срок окупаемости находится в диапазоне 0,63-0,83 года. Однолетний ROI относительно исходных инвестиций составляет 20-60%, а недисконтированный пятилетний ROI - 500-700%. При ставке 12% пятилетний NPV находится примерно в диапазоне 831-1192 млн долл. без учета дополнительных эксплуатационных затрат, которые в источнике отдельно не раскрыты [21].

Показатель UPS ORION	Значение	Комментарий
CapEx / build-deploy cost	250 млн долл.	INFORMS [21]
Годовая экономия, lower bound	300 млн долл./год	INFORMS [21]
Годовая экономия, upper bound	400 млн долл./год	INFORMS [21]
Простой payback	0,63-0,83 года	250/400; 250/300
ROI за 1 год	20-60%	(B - 250) / 250
NPV 5 лет при r = 12%	831-1192 млн долл.	расчет по формуле (3)

Таблица 6. Реальный ROI-кейс процессной AI-оптимизации на примере UPS ORION.

Риск-экономический компонент может быть представлен через IMA-сценарий из авторской работы. При SLE 4,88 млн долл., ARO 0,1 и коэффициенте снижения риска 50% предотвращаемый ущерб составляет 244 тыс. долл. в год. При InitialEx 150 тыс. долл. и OpEx 180 тыс. долл. в год чистый годовой поток составляет 64 тыс. долл.; пятилетний NPV при r = 12% равен около 80,7 тыс. долл., PI около 1,54, DPP около 2,9 года [2]. Этот расчет демонстрирует, что governance и безопасность не являются только затратами: при значимом ALE они формируют самостоятельный источник положительного NPV.

Показатель IMA	Значение	Экономическая интерпретация
InitialEx	150 000 долл.	развертывание и интеграция IMA-контура [2]
OpEx	180 000 долл./год	инференс, инфраструктура и MLOps/SecOps [2]
Предотвращаемые потери	244 000 долл./год	SLE 4,88 млн × ARO 0,1 × k 50% [2]
Чистый CF	64 000 долл./год	244 000 - 180 000
NPV, r = 12%, N = 5	80 700 долл.	расчет DCF в авторской статье [2]
PI / DPP	PI ≈ 1,54; DPP ≈ 2,9 года	инвестиционная целесообразность [2]

Таблица 7. Риск-ориентированная экономика IMA как пример расчета ALE/NPV для агентных систем.

## 7. Как такая экономика переносится на Century без фактических пилотных данных

Фактические логи, коммерческие чеки и метрики пилотов Century отсутствуют. Поэтому корректная экономическая часть не должна содержать утверждений о достигнутом снижении затрат или фактическом ROI Century. Вместо этого может быть построена нормативная модель будущего пилота, в которой реальные внешние цифры используются для задания диапазона ожиданий, а результат Century определяется только после измерения baseline/after.

Пилот Century целесообразно проектировать не как демонстрацию ассистента, а как controlled process experiment. Минимальный набор данных включает: число кейсов, среднее и медианное время обработки, SLA-просрочки, repeat rate, долю ручных эскалаций, стоимость минуты операционного времени, стоимость контроля, стоимость инференса, долю ответов с валидными источниками, число ошибок и риск-событий. Для каждого процесса рассчитываются E\_net, ROI, NPV и DPP; внешние бенчмарки используются только для предварительной оценки power и ожидаемого диапазона эффекта.

Для клиентского сервиса с масштабом 100 тыс. обращений в год нормализованная таблица 3 показывает, что переход от ручного baseline 11 минут к агентному уровню 2 минуты дает около 15 тыс. часов высвобождения. При монетизируемой стоимости 30 долл./ч это 450 тыс. долл. временного эффекта до учета платформенных затрат, повторных обращений и риска. Если all-in стоимость такого сценария составляет 100 тыс. долл. в год, ROI только по времени равен 350%; при стоимости 250 тыс. долл. - 80%; при стоимости 500 тыс. долл. - около -10%. Следовательно, достаточный масштаб процесса и переиспользование платформенных компонентов являются критическими условиями положительного ROI.

All-in cost сценария, долл./год	Benefit по времени при \$30/ч	ROI только по времени	Комментарий
100 000	450 000	350%	Без учета -25% repeat inquiries и риск-эффекта [16]
250 000	450 000	80%	Без учета -25% repeat inquiries и риск-эффекта [16]
500 000	450 000	-10%	Без учета -25% repeat inquiries и риск-эффекта [16]

Таблица 8. Сценарная экономика Century-like customer-support workflow на 100 тыс. обращений.

При портфельном внедрении экономика улучшается за счет reuse-эффекта. Один и тот же контур RBAC, RAG-индексации, аудита, метрик, обработки документов и интеграции может использоваться в нескольких процессах: клиентская поддержка, договоры, закупки, регламенты, комплаенс, внутренние запросы, финансовые документы. Если первый сценарий несет высокую долю интеграционных затрат, следующий сценарий должен иметь меньший  $C_{int}$  и  $C_{change}$ . Именно этот эффект превращает Century из инструмента в платформу.

$$C_{int, p+1} = C_{int, p} \cdot (1 - \lambda_{reuse}), \quad E_{portfolio} = \sum_{p=1}^n E_{net, p} - C_{shared}$$

(5) портфельная экономика платформы за счет переиспользования сервисов

Формула (5) фиксирует платформенную логику. Если  $\lambda_{reuse}$  высок, предельная стоимость нового workflow снижается. Это важнее для инвестора, чем одиночный ROI пилота, поскольку рыночная ценность Century зависит от расширения числа процессов у клиента, а не от одного ассистента. В пилотной отчетности по Century должны фиксироваться reuse rate, marginal deployment cost, cost per agentic run и доля общих сервисов в новом сценарии.

## 8. Сценарии внедрения Century и инвестиционная логика

Сценарий 1 - пилот одного процесса. Целью является не максимальная автономность, а доказательство измеримого эффекта. Оптимальная длительность составляет 8-12 недель. Выбирается процесс с повторяемыми кейсами, формализованным регламентом, документной базой и понятными KPI: обработка обращений, подготовка ответа по регламенту, анализ договорного пакета, закупочная проверка, комплаенс-опросник, внутренний service desk. Критерии завершения: baseline до внедрения, after-период, контроль case mix, журнал решений, доля HITL, ошибки, стоимость обработки и расчет ROI.

Сценарий 2 - портфель 3-5 процессов. Целью является доказательство платформенной экономики. В этом режиме Century должна показать не только эффект отдельного workflow, но и снижение предельной стоимости следующего сценария. Экономическими метриками становятся reuse rate, cost per case, cost per workflow launch, SLA improvement, доля ответов с проверенными источниками, доля ручных эскалаций и динамика затрат на контроль.

Сценарий 3 - enterprise deployment. Платформа становится элементом операционной инфраструктуры: единый контур документов, знаний, workflow, ролей, аудита, метрик и управления стоимостью. Для клиента ценность выражается в снижении времени обработки, repeat rate, дефектов, SLA-просрочек и риска. Для инвестора ценность выражается в expansion revenue, retention, увеличении числа активных сценариев на клиента, снижении marginal cost внедрения и накоплении evidence base с анонимизированными метриками.

Сценарий	Масштаб	Основные метрики	Инвестиционный смысл
Pilot	1 процесс; 8-12 недель	T_before/T_after, error rate, repeat rate, HITL share, cost per case	Go/no-go по фактическому ROI и риску
Portfolio	3-5 процессов	Reuse rate, marginal deployment cost, SLA, throughput, cost observability	Доказательство платформенной экономики
Enterprise	Единый контур AI-workflow	NPV, PI, DPP, retention, expansion revenue, risk-adjusted savings	Инвестиционная масштабируемость

Таблица 9. Сценарии внедрения Century и показатели инвестиционной проверки.

Инвестиционная логика Century должна быть уже, чем общий рынок AI. Макропрогнозы Gartner и IDC полезны для подтверждения направления капитальных потоков, но не формируют TAM платформы автоматически [7; 8]. Более точная гипотеза: Century может претендовать на сегмент enterprise agentic workflow governance для организаций, где критичны документы, источники, роли, локальное или изолированное развертывание, аудит и подтверждаемая экономия. В таком сегменте барьером конкуренции становится не только модель, а интеграция с процессом и доказательная дисциплина.

Сравнение с глобальными платформами следует проводить по слоям. Microsoft Copilot Studio силен в Microsoft-экосистеме и controlled autonomous agents [23]. Salesforce Agentforce ориентирован на CRM, sales и service [24]. ServiceNow AI Agents сильны внутри IT/HR/CSM workflows [25]. UiPath развивает agentic automation на базе RPA и оркестрации людей, роботов и агентов [26]. Celonis AgentC усиливает process intelligence и process-aware agents [27]. SAP Joule и IBM watsonx Orchestrate работают как embedded enterprise agents в крупных suite-экосистемах [28; 29]. Century рационально позиционировать как локально разворачиваемый, audit-ready, knowledge/workflow-oriented контур для организаций, где vendor lock-in и требования к данным являются существенными ограничениями.

## 9. Риски, ограничения и требования к доказательству эффекта

Первое ограничение связано с источниками. Денежные эффекты Klarna, С.Н. Robinson, Accenture/Celonis, Microsoft и UPS относятся к разным типам процессов, разной методологической надежности и разному уровню раскрытия данных. Vendor/customer cases с конкретными цифрами полезны для экономической калибровки, но без raw logs и контрольной группы не доказывают причинный эффект в строгом научном смысле. Поэтому в работе они используются как бенчмарки и пороги, а не как результаты Century.

Второе ограничение связано с безопасностью. Agentic systems имеют специфические риски: prompt injection, sensitive information disclosure, data poisoning, improper output handling, excessive agency, system prompt leakage, vector and embedding weaknesses и unbounded consumption. OWASP Top 10 for LLM Applications 2025 указывает эти категории как ключевые для LLM-приложений [30]. Для Century-like платформы это означает необходимость проектировать RBAC, изоляцию инструментов, контроль источников, red teaming, RAG trust, журналирование и безопасный отказ как финансово значимые элементы, а не как постфактум-дополнение.

Третье ограничение связано с организационной монетизацией. Высвобождение часов становится денежным потоком только при сокращении внешних затрат, перераспределении штата на ценностные задачи, росте обслуживаемого объема без расширения FTE, снижении аутсорсинга или повышении выручки. Если организация просто добавляет AI поверх прежнего процесса, ROI может оказаться отрицательным даже при положительной технической метрике. Поэтому в пилоте Century обязательны финансовый владелец, владелец процесса и заранее определенное правило монетизации.

Для доказательства эффекта Century должна формировать evidence pipeline. Каждый production-сценарий должен завершаться анонимизированной карточкой: процесс, объем кейсов, baseline, after, case mix, уровень автономности, доля HITL, cost per case, ошибки, repeat rate, SLA, риск-события, стоимость инференса, операционные затраты, чистый эффект и NPV. Именно такой массив данных превратит платформу из перспективной архитектуры в инвестиционно проверяемый актив.

## Заключение

ИИ-агентские платформы создают измеримый экономический эффект только тогда, когда становятся частью управляемого процессного контура. Одиночный LLM ускоряет отдельную операцию, но не гарантирует ROI. Агентская платформа соединяет данные, документы, RAG, workflow, RBAC, Human-in-the-Loop, audit trail и KPI, превращая AI в управляемую инфраструктуру производительности и риска.

Century соответствует архитектурным условиям такой платформы: публичное описание фиксирует документы и знания, workflow, контроль доступа, аудит, метрики и локальные/изолированные/гибридные режимы развертывания [4-6]. Отсутствие фактических логов и коммерческих параметров не позволяет заявлять собственный фактический ROI Century, но позволяет построить корректную экономическую модель будущих внедрений. На основе реальных внешних

кейсов показано, что AI-эффект может выражаться в десятках миллионов долларов годовой прибыли или working-capital benefits в customer support и procurement, сотнях миллионов долларов экономии в route optimization, а также положительном NPV риск-контроля при значимом ALE [2; 16; 18; 21; 22].

Научный вклад работы состоит в предложении attribution-safe ROI: внешние результаты используются не как рекламное доказательство, а как калибровочные бенчмарки, переносимые через коэффициенты готовности, сходства, монетизации и governance. Практический вклад заключается в расчетных таблицах, нормализованном сравнении без AI, LLM-assist и agentic workflow, порогах окупаемости и сценариях внедрения Century. Следующий шаг для платформы - пилот с фиксированным baseline, контролем case mix, журналом решений, DCF-моделью и публично пригодной карточкой эффекта.

## Список источников

- [1] Пискун Е., Крячев Е., Азизов А. Риск-ориентированная оценка экономической эффективности LLM в сервисном секторе Республики Беларусь // Наука и инновации. 2026. №3. С. 36-44. DOI: 10.29235/1818-9857-2026-03-36-44. URL: <https://innosfera.belnauka.by/jour/article/view/1001>.
- [2] Пискун Е. С., Азизов А. А., Крячев Е. В. Методика оценки финансовых рисков организаций на основе внедрения Isolated Multiagent Arbitration // Цифровая трансформация. 2026. URL: <https://dt.bsuir.by/jour/article/view/1005>.
- [3] Пискун Е. С., Азизов А. А., Крячев Е. В. Управляемая интеграция AI-агентов в бизнес-процессы организации: архитектура, методика и риск-ориентированная оценка эффекта. Приложенный материал ai\_agents\_article\_formatted\_final.docx, 2026.
- [4] Century. Платформа управляемого внедрения корпоративного искусственного интеллекта. URL: <https://century-ai.ru/> (дата доступа: 07.06.2026).
- [5] Stacklevel Group. Century: управляемое LLM-развертывание. URL: <https://genai.by/ru/products/century/> (дата доступа: 07.06.2026).
- [6] Century. Кейсы и сценарии применения. URL: <https://century-ai.ru/cases> (дата доступа: 07.06.2026).
- [7] Gartner. Gartner Says Worldwide AI Spending Will Total \$1.5 Trillion in 2025. 2025. URL: <https://www.gartner.com/en/newsroom/press-releases/2025-09-17-gartner-says-worldwide-ai-spending-will-total-1-point-5-trillion-in-2025>.
- [8] IDC / BusinessWire. Worldwide Spending on Artificial Intelligence Forecast to Reach \$632 Billion in 2028. 2024. URL: <https://www.businesswire.com/news/home/20240819177906/en/Worldwide-Spending-on-Artificial-Intelligence-Forecast-to-Reach-%24632-Billion-in-2028-According-to-a-New-IDC-Spending-Guide>.
- [9] McKinsey & Company. The economic potential of generative AI: The next productivity frontier. 2023. URL: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>.
- [10] McKinsey & Company. The state of AI in 2025: Agents, innovation, and transformation. 2025. URL: [https://www.mckinsey.com/~media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/november%202025/the-state-of-ai-2025-agents-innovation\\_cmyk-v1.pdf](https://www.mckinsey.com/~media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/november%202025/the-state-of-ai-2025-agents-innovation_cmyk-v1.pdf).
- [11] Brynjolfsson E., Li D., Raymond L. Generative AI at Work. NBER Working Paper No. 31161. 2023; QJE version. URL: <https://www.nber.org/papers/w31161>.
- [12] Noy S., Zhang W. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence // Science. 2023. DOI: 10.1126/science.adh2586. URL: <https://www.science.org/doi/10.1126/science.adh2586>.
- [13] Peng S., Kalliamvakou E., Cihon P., Demirer M. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. 2023. URL: <https://arxiv.org/abs/2302.06590>.
- [14] Microsoft. Randomized Controlled Trial for Microsoft Copilot for Security. Whitepaper. 2024. URL: <https://www.microsoft.com/content/dam/microsoft/final/en-us/microsoft-product-and-services/microsoft-dynamics-365/pdf/Microsoft-Copilot-for-Security-productivity-findings-Whitepaper-Jan2024.pdf>.
- [15] Dell'Acqua F. et al. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School, 2023. URL: <https://www.hbs.edu/faculty/Pages/item.aspx?num=64700>.
- [16] Klarna. Klarna AI assistant handles two-thirds of customer service chats in its first month. 2024. URL: <https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/>; OpenAI. Klarna's AI assistant does the work of 700 full-time agents. URL: <https://openai.com/index/klarna/>.
- [17] C.H. Robinson. Using Generative AI, C.H. Robinson Has Achieved Automation Across the Entire Lifecycle of a Freight Shipment. 2024. URL: <https://www.chrobinson.com/en-us/about-us/newsroom/press-releases/2024/generative-ai-for-freight-shipment-lifecycle/>.
- [18] Celonis. Accenture + Celonis: Process Mining. URL: <https://www.celonis.com/solutions/stories/accenture-purchase-to-pay-process-mining>.
- [19] Accenture. Mining a Procurement Function. URL: <https://www.accenture.com/us-en/case-studies/about/turning-process-friction-into-flow>.
- [20] UiPath. Isbank's Journey of Operational Excellence with Process Mining. URL: <https://www.uipath.com/resources/automation-case-studies/isbank-journey-of-operational-excellence-with-process-mining>.
- [21] INFORMS. UPS ORION Project. URL: <https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/UPS>.
- [22] Reuters / Yahoo Finance. Microsoft racks up over \$500 million in AI savings while slashing jobs, Bloomberg News reports. 2025. URL: <https://finance.yahoo.com/news/microsoft-racks-over-500-million-184531782.html>.
- [23] Microsoft Learn. Design autonomous agent capabilities in Microsoft Copilot Studio. URL: <https://learn.microsoft.com/en-us/microsoft-copilot-studio/guidance/autonomous-agents>.

- [24] Salesforce. Agentforce: The AI Agent Platform. URL: <https://www.salesforce.com/agentforce/>.
- [25] ServiceNow. AI Agents. URL: <https://www.servicenow.com/products/ai-agents.html>.
- [26] UiPath. UiPath launches enterprise-grade platform for agentic automation. URL: <https://www.uipath.com/newsroom/uipath-launches-first-enterprise-grade-platform-for-agentic-automation>.
- [27] Celonis. Celonis Showcases Latest Process Intelligence and AI Innovations at Celonis:Next 2025. URL: <https://www.celonis.com/news/press/celonis-showcases-latest-process-intelligence-and-ai-innovations-at-next-2025>.
- [28] SAP News Center. SAP Connect: new Joule agents and embedded intelligence. 2025. URL: <https://news.sap.com/2025/10/sap-connect-business-ai-new-joule-agents-embedded-intelligence/>.
- [29] IBM. watsonx Orchestrate. URL: <https://www.ibm.com/products/watsonx-orchestrate>.
- [30] OWASP. Top 10 for Large Language Model Applications 2025. URL: <https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf>.
- [31] Meimandi K. J. et al. The Measurement Imbalance in Agentic AI Evaluation Undermines Industry Productivity Claims. arXiv:2506.02064, 2025. URL: <https://arxiv.org/abs/2506.02064>.