

ПРЕДЕЛ АВТОНОМНОСТИ ИИ-АГЕНТА В КРЕДИТНЫХ ОРГАНИЗАЦИЯХ: ЭКОНОМИЧЕСКИЙ ПАРАДОКС СНИЖЕНИЯ ИЗДЕРЖЕК И РОСТА ПОТЕРЬ

Т. Н. Зверькова

Оренбургский государственный университет,
Оренбург, Россия

Аннотация. Настоящая статья посвящена вопросам применения искусственного интеллекта (ИИ) в кредитных организациях. Цель исследования состоит в обосновании предельного уровня участия ИИ-агента в банковской деятельности, при котором дальнейшее сокращение роли работника (специалиста банка) перестаёт улучшать экономическую отдачу применения агента. Методологическая основа исследования включает математическое моделирование зависимости между степенью участия ИИ-агента, положительным результатом и затратами, возникающими при повторяемой ошибке. Результаты работы. Показано, что расширение роли ИИ-агента снижает текущие затраты и ускоряет прохождение стандартных банковских операций, однако одновременно увеличивает цену ошибочной логики, если она воспроизводится в потоке однородных действий до момента выявления. Обосновано, что выгода от передачи решений агенту растёт с убывающим приростом, тогда как затраты возрастают по мере приближения к полной автономии. Выводы. Полная передача выполнения операций ИИ-агенту не должна рассматриваться как заранее предпочтительный вариант для банка. Экономически оправданным является такой уровень участия агента, при котором прирост дохода превышает рост затрат, связанных с повторяемой ошибкой.

Ключевые слова: ИИ-агент, банковские операции, повторяемая ошибка, ожидаемые потери.

Цитирование: Зверькова Т. Н. Предел автономности ИИ-агента в кредитных организациях: экономический парадокс снижения издержек и роста потерь // Сибирская финансовая школа. 2026. № 2 (162). С. 13-20. DOI: 10.34020/1993-4386-2026-2-13-20

Введение

На раннем этапе цифровизация банковского дела была связана главным образом с удалённым обслуживанием, электронным документооборотом, автоматизацией расчётов и сокращением ручной обработки. Эти изменения затрагивали, прежде всего, способ взаимодействия с клиентом и движение информации внутри кредитной организации, но не меняли сам порядок оценки, подтверждения и исполнения операций. Использование ИИ-агентов меняет эту ситуацию, поскольку агент уже не ограничивается ускорением передачи данных, а участвует в цифровой трансформации банковской деятельности, то есть в проверке сведений, выборе дальнейшего действия и внедрении процессов. В результате возникает новая проблема, связанная с повторением ошибки в операционном потоке. Если в настройку агента попадает неверное правило, оно применяется не к одному случаю, а к группе однотипных действий до момента выявления отклонения. Поэтому оценка ИИ-агента не может сводиться

к сокращению текущих затрат и времени обработки, поскольку необходимо учитывать возможные потери от серийного воспроизведения ошибочной логики. Для кредитной организации такая проблема особенно чувствительна, поскольку ущерб может проявиться позднее через ухудшение кредитного портфеля, претензии клиентов или дополнительные проверки.

Цель статьи состоит в теоретическом обосновании предельного уровня участия ИИ-агента в банковской деятельности, при котором дальнейшее сокращение роли сотрудника перестаёт улучшать экономическую оценку применения агента.

Для достижения поставленной цели решаются две задачи. Первая задача состоит в определении того, почему роль ИИ-агента нельзя оценивать только через ускорение обработки и снижение расходов. Вторая - заключается в построении зависимости между объёмом действий, передаваемых ИИ, и возникающими потерями, и в доказательстве условий, при которых дальнейший рост участия агента перестаёт улучшать результат.

Обзор исследований по теме

Современные зарубежные исследования применения искусственного интеллекта всё чаще обращаются к теме использования ИИ-агентов в кредитных организациях. В таких работах обычно рассматриваются преимущества агентов, повышение скорости обработки данных, снижение текущих затрат и прозрачность формируемого вывода. Это согласуется с природой банковской деятельности, поскольку последствия многих операций проявляются не сразу. D. Diamond трактует финансового посредника как институт мониторинга заёмщиков, снижающего издержки индивидуального наблюдения за их поведением [1]. D. Diamond и P. Dybvig показывают, что преобразование краткосрочных обязательств в долгосрочные активы переносит последствия текущей операции в будущие периоды [2]. D. Andolfatto и его соавторы объясняют нестабильность посредников не случайными сбоями, а особенностями их деятельности [3]. P. McLeMore, A. Mihov и L. Sanz анализируют использование ИИ через призму операционных потерь и показывают, что при расширении цифровых процедур ущерб может возрасти из-за сбоев, ошибок и зависимости от сторонних поставщиков [4]. T. Ciuraga описывает ИИ не только как средство ускорения обработки, но и как фактор, меняющий последующую организацию процессов после внедрения [5]. Эти работы позволяют перейти к анализу расходов, возникающих при регулярном применении ИИ в банковской практике.

В литературе о машинном обучении основное внимание уделяется верности прогноза, качеству данных, объяснимости и сопоставлению методов классификации. S. Zhang и L. Yu систематизируют такие методы применительно к кредитной оценке [6], M. M. Haque Muki, F. Hasan и T. Choudhury подчеркивают, что исследования в этой сфере сосредоточены на объяснимости, качестве данных и сравнении алгоритмов [7], а A. Awadelkarim отмечает высокий уровень прогнозирования дефолта на исторических данных [8]. R. Bahloul, N. Hewahi и W. Elmedany дополняют эту дискуссию вопросами справедливости и объяснимости [9], тогда как H. Sadok, F. Sakka и M. Maknoui связывают применение ИИ не только с прогнозом дефолта, но и с изменением порядка оценки, обработки данных и последующего контроля [10]. В совокупности эти публикации показывают, что оценка ИИ в банке не может ограничиваться точностью на проверочном массиве, поскольку после внедрения большее значение приобретает устойчивость работы агента.

Исследования агентных систем позволяют точнее раскрыть эту проблему. J. Danielsson и A. Uthmanн показывают, что ИИ может усиливать уязвимость финансовой системы, если алгоритмы используют похожие источники данных и начинают действовать аналогично [11]. A. Bernardelli и его партнеры подчеркивают зависимость результата применения ИИ от качества данных, настройки и области использования [12]. S. Ganesh со своими

коллегами описывает агентов как системы, которые работают с разными источниками информации, выполняют последовательные действия и взаимодействуют с внешней цифровой средой [13]. M. McClellan анализирует риск согласованного поведения при широком применении генеративного ИИ [14]. M. Mathen и A. Paul рассматривают ответственное применение ИИ через справедливость, объяснимость и операционные ограничения [15]. A. Skouloudis и A. Venkatraman отмечают, что внедрение ИИ связано не только с техническими трудностями, но и с организационными барьерами, конфликтами интересов и распределением ответственности [16].

В этой связи недостаточно разработанным остаётся способ обоснования роли ИИ-агента в банковских процессах. Существующая литература показывает отдельные свойства ИИ-систем, но не даёт расчётного критерия, позволяющего определить, при каком объёме участия агента текущая экономия ещё соразмерна последующим затратам на проверку, повторный разбор спорных случаев и исправление ранее выполненных действий. Поэтому в статье участие ИИ-агента описывается через соотношение между снижением затрат в момент выполнения операции и расходами, которые могут возникнуть после её завершения.

Основная часть

Банк отличается от обычной организации тем, что финансовые последствия его операций нередко обнаруживаются уже после их выполнения и зависят от дальнейшего поведения заёмщика, клиента или контрагента. В банковской деятельности характерно наложение текущих затрат на будущие последствия, поскольку первоначально выгодная операция позднее может привести к просрочке, претензиям, дополнительным проверкам или расходам на исправление. Поэтому для кредитной организации значение имеет не сам факт применения ИИ-агента, а объём действий, выполняемых без предварительного подтверждения специалистом банка. Когда агент только готовит сведения для специалиста, ошибка обычно не выходит за пределы отдельного случая. Если цифровой исполнитель начинает проводить типовые операции без его участия, неверная логика способна повторяться в серии однородных действий. Следовательно, оценка применения ИИ-агента должна начинаться с разграничения уровней его участия в процессах.

Для этого вводится формальная конструкция, позволяющая описать, какая часть операции выполняется агентом самостоятельно, а какая остаётся за работником кредитной организации. Введём переменную α , где $\alpha \in [0,1]$, характеризующую степень самостоятельного выполнения действий ИИ-агентом. При $\alpha = 0$ агент не выполняет их без работника банка и используется только для подготовки сведений, справок или предварительного варианта ответа. При $\alpha = 1$ агент автономно работает в пределах заданного порядка. Между этими

крайними состояниями располагаются промежуточные режимы, при которых стандартные случаи обрабатываются агентом, а спорные направляются специалисту банка.

Обозначим через $E(\alpha)$ положительную экономическую величину, возникающую при увеличении объёма процессов, поручаемых агенту. Она связана с сокращением затрат, ускорением рассмотрения заявок, обработкой платежей, проверкой сведений и разбором типовых обращений. Зависимость $E(\alpha)$ является возрастающей, поскольку при увеличении α агент принимает на себя больший объём стандартных действий $E'(\alpha) > 0$. При этом предельный прирост $E(\alpha)$ уменьшается, поскольку сначала агенту поручаются наиболее типовые действия, а затем остаются спорные случаи, требующие проверки, пояснения и возможного пересмотра: $E''(\alpha) < 0$. Дополнительно зададим условия, при которых положительный эффект начинается с нулевого уровня и возрастает на начальном участке: $E(0) = 0$; $E'(0) > 0$.

Обозначим через $C(\alpha)$ стоимость риска, контроля и последующего исправления. Она включает затраты на выполнение, сопровождение, проверку данных, разбор нестандартных случаев, стоимость привлечённых ресурсов, ожидаемые потери, нагрузку на капитал, внутренние проверки и исправление последствий после выявления ошибки. Данная зависимость также возрастает, поскольку увеличивается поток итераций, проходящих через одну и ту же логику агента: $C'(\alpha) > 0$. В отличие от $E(\alpha)$, стоимость риска растёт с предельным приростом, поэтому $C''(\alpha) > 0$.

Для обоснования внутреннего оптимума не нужно предполагать, что стоимость риска при полном поручении функций агенту уходит к бесконечности. Достаточно зафиксировать изменение соотношения на границах интервала: на начальном участке прирост положительной величины превышает увеличение риск-затрат, а при почти полном участии агента дополнительные потери оказываются выше дальнейшей экономии: $E'(0) > C'(0)$; $E'(1) < C'(1)$.

Совокупную оценку применения ИИ-агента обозначим через $V(\alpha)$. Она определяется как разность: $V(\alpha) = E(\alpha) - C(\alpha)$. Задача состоит в нахождении такого значения α , при котором $V(\alpha)$ достигает максимума. Поэтому расчёт направлен на выявление точки, после которой новая передача функций уже не покрывает возрастающую цену ошибочной логики.

Наилучшее значение $V(\alpha)$ находится там, где прирост положительной величины равен увеличению затрат и потерь: $E'(\alpha^*) = C'(\alpha^*)$. Такое значение α^* существует внутри интервала $[0,1]$. Это следует из того, что в начале разность $E'(\alpha) - C'(\alpha)$ положительна, а у правой границы становится отрицательной. Кроме того, эта разность убывает на всём интервале, поскольку $E'(\alpha)$ уменьшается, а $C'(\alpha)$ увеличивается. Поэтому точка α^* является единственной.

Вторая производная $V(\alpha)$ имеет вид: $V''(\alpha) = E''(\alpha) - C''(\alpha)$. Поскольку $E''(\alpha) < 0$, а $C''(\alpha) > 0$, получаем $V''(\alpha) < 0$ на всём отрезке. Следовательно, $V(\alpha)$ достигает единственного максимума в точке α^* . Это означает, что рациональная доля действий, выполняемых ИИ-агентом, находится внутри интервала и не совпадает с полным переходом процедуры к цифровому исполнителю.

В обычном процессе ошибка специалиста чаще связана с отдельной заявкой, платежом или проверяемым случаем. После полученного замечания, нового документа или внутренней проверки работник банка может изменить дальнейшее действие. ИИ-агент воспроизводит однотипную логику в значительном количестве итераций до внесения изменений, ограничения его полномочий или переобучения. Именно поэтому повторяемость ошибки получает самостоятельное значение.

Для описания этого механизма обозначим через N число случаев соответствующего типа за выбранный период. Пусть p показывает вероятность появления неверной логики, L выражает потерю по одному случаю, $q(\alpha)$ отражает долю потока, охваченную таким просчетом при заданном значении α , а $\tau(\alpha)$ отражает длительность скрытого накопления последствий. Тогда ожидаемые потери от повторяемой ошибки можно представить следующим образом: $C_{repeat}(\alpha) = p \times L \times N \times q(\alpha) \times \tau(\alpha)$.

Эта формула показывает, почему потери могут увеличиваться быстрее при расширении роли ИИ-агента. С ростом α больше решений проходит через одну и ту же агентную логику, поэтому увеличивается $q(\alpha)$. Одновременно может возрастать $\tau(\alpha)$, поскольку спорные случаи реже попадают на проверку. В результате потери определяются не одним неверным решением, а произведением размера потока, охвата ошибки и периода её скрытого повторения.

Для отражения временного измерения качества агента можно ввести показатель $K(t)$, который характеризует соответствие текущих итераций агента уровню, принятому при внедрении или последней настройке. Если $K(t)$ близок к единице, агент сохраняет проверенные ранее параметры работы. Снижение $K(t)$ означает рост пересмотров, исправлений, жалоб, пропущенных отклонений и расхождений между ожидаемыми и фактическими потерями. Такой показатель должен опираться не на правильность ответа, а на наблюдаемые признаки: для кредитных операций это доля повторных рассмотрений, для расчётных процедур – возвраты и претензии, для внутреннего контроля – позднее выявление нарушений.

С учётом временного фактора совокупная оценка применения агента может быть записана как $V(\alpha, t) = E(\alpha)K(t) - C(\alpha, t)$, где $C(\alpha, t)$ включает не только затраты, заданные в исходной постановке, но и расходы, возникающие при ухудшении работы агента после внедрения. Отдельное значение имеет задержка между первым сигналом ухуд-

шения и фактическим вмешательством кредитной организации. Такие затраты можно представить как:

$C_{delay} = \int_{t_s}^{t_r} [B_{max} - B(t)]dt$, где t_s обозначает момент появления сигнала об ухудшении; t_r указывает момент вмешательства; B_{max} показывает возможный итог при своевременной корректировке; $B(t)$ отражает фактическую величину при продолжающемся ухудшении.

Для целей настоящей статьи достаточно статической постановки, поскольку задача состоит в доказательстве существования внутреннего значения α^* и пояснении базовой формы зависимости между участием агента, положительной величиной и затратами от повторяемой ошибки. Поэтому в наглядной демонстрации предполагается, что $K(t) = 1$, задержка вмешательства отсутствует, а совокупная оценка задаётся выражением $V(\alpha) = E(\alpha) - C(\alpha)$. Временные изменения работы агента, деградация параметров и запаздывание реакции могут рассматриваться как отдельное направление дальнейшей проверки предложенной конструкции. Демонстрация не является расчётом для конкретной кредитной организации и служит только для показа формы зависимости между α , $E(\alpha)$, $C(\alpha)$ и $V(\alpha)$.

Положительный итог определяется как: $E(\alpha) = 100(1 - e^{-3\alpha})$. Затраты и потери от повторяемой ошибочной логики: $C(\alpha) = 10 \frac{\alpha^2}{(1 - \alpha + 0,1)}$.¹ $E(\alpha)$

быстро возрастает на начальном участке, после чего её прирост постепенно уменьшается. $C(\alpha)$, напротив, сначала увеличивается умеренно, а затем ускоряется при приближении к полному поручению

банковских действий агенту. При этом $C(\alpha)$ не стремится к бесконечности при $\alpha = 1$, поскольку знаменатель остаётся положительным. Поэтому демонстрация показывает не бесконечный рост потерь, а резкое увеличение затрат около правой границы интервала. (см.табл.1).

Данные таблицы показывают, что $V(\alpha)$ растёт до значения $\alpha = 0,6$, после чего начинает снижаться. При $\alpha = 0,6$ положительный эффект ещё существенно превышает затраты и потери от повторяемой ошибки. При более точном расчёте максимум находится внутри интервала между 0,6 и 0,7. При $\alpha = 0,8$ рост $C(\alpha)$ уже заметно уменьшает совокупную оценку применения агента. При $\alpha = 1,0$ затраты и потери в выбранной демонстрации превышают положительную величину от полного поручения действий ИИ-агенту (см. рис.1). Причина состоит в механизме повторяемой ошибки: чем ближе α к единице, тем меньше остаётся точек человеческого контроля, способных остановить ошибочную логику на ранней стадии. В результате одна и та же неверная последовательность действий агента может охватить почти весь поток однотипных операций, а период её скрытого сохранения становится максимальным. Поэтому даже небольшое увеличение самостоятельности агента на участке от 0,9 до 1,0 способно привести к многократному росту ожидаемых потерь, тогда как выгода увеличивается лишь незначительно. Результат $V(\alpha)$ в такой ситуации резко снижается и переходит в отрицательную область. Это показывает, что полное выполнение операции ИИ-агентом экономически не оправдано, а оптимальный уровень его самостоятельности должен находиться ниже единицы.

Таблица 1

Значения $E(\alpha)$, $C(\alpha)$, $V(\alpha)$ при разных уровнях α^*

α	$E(\alpha)$	$C(\alpha)$	$V(\alpha)$
0	0	0	0
0,1	25,9	0,1	25,8
0,2	45,1	0,4	44,7
0,3	59,3	1,1	58,2
0,4	69,9	2,3	67,6
0,5	77,7	4,2	73,5
0,6	83,5	7,2	76,3
0,7	87,8	12,2	75,5
0,8	90,9	21,3	69,6
0,9	93,3	40,5	52,8
1	95	100	-5

*Источник: рассчитано автором.

¹ Приведённые числовые параметры выбраны исключительно для демонстрации качественного поведения функций и не претендуют на эмпирическую точность. В реальном применении они должны быть оценены на основе данных конкретной кредитной организации.

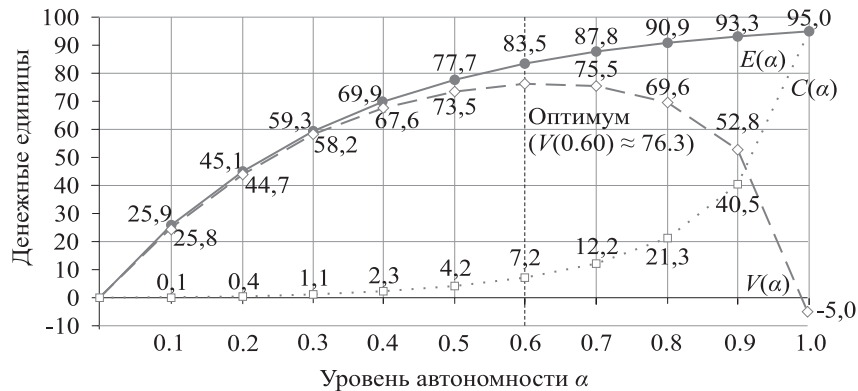


Рис. 1. Оптимальное значение α при применении ИИ-агента в банке

Для практического применения параметры зависимостей должны определяться по данным конкретной кредитной организации. $E(\alpha)$ может оцениваться через снижение затрат, сокращение времени рассмотрения заявок, ускорение обработки платежей и уменьшение числа повторных проверок. $C(\alpha)$ должна учитывать ожидаемые потери, нагрузку на капитал, расходы на внутренние проверки, исправление ошибок, инциденты и прекращение использования соответствующего сценария. Поэтому демонстрация имеет вспомогательное значение и служит только для пояснения формы зависимости.

Доля действий, поручаемых ИИ-агенту, не должна становиться самостоятельной целью цифрового проекта. Кредитная организация может расширять роль агента только до той точки, где доход ещё превышает возможные потери от повторяемой ошибочной логики. Для простых и хорошо стандартизированных действий допустимое значение α может быть сравнительно высоким, поскольку цена ошибки и срок её обнаружения остаются управляемыми. Для кредитования, внутреннего контроля и клиентских споров этот показатель должен быть ниже, поскольку неблагоприятные последствия проявляются позднее и требуют более дорогого исправления. Поэтому единый уровень участия ИИ-агента для всех процедур был бы методически неверным.

Место специалиста банка в процессе с участием агента также меняется, но не исчезает полностью. Специалист не обязательно остаётся основным исполнителем, однако его роль смещается к проверке пограничных случаев, разбору отклонений, остановке повторяемой ошибки и подтверждению операций с высокой ценой последствий. Поэтому вопрос состоит не в полном исключении человека, а в разумном распределении действий между агентом и специалистом. Такое разделение должно определяться ценой ошибки, скоростью её выявления и возможностью исправить последствия без накопления потерь.

Качество работы после внедрения агента требует постоянного наблюдения, поскольку исходная проверка не гарантирует сохранения прежнего уровня. Чем больше операций проходит через агента, тем короче должен быть период между первым признаком ухудшения и вмешательством кредитной организации. При запоздалой реакции цена ошибки быстро увеличивается, поскольку неверная логика успевает пройти через большой поток однородных случаев. Поэтому расширение роли ИИ-агента требует не только точной настройки, но и регулярного контроля качества.

Инвестиционная оценка проекта также должна учитывать специфику применения агента. ROI, NPV и срок возврата вложений нельзя рассчитывать только на основании сокращения затрат и ускорения процедур. В расчёт должны входить значение α , изменение качества работы во времени и потери от повторяемой ошибки. Иначе проект может выглядеть привлекательным по техническим показателям, но терять смысл после учёта последствий.

Предложенный подход применим не ко всем случаям использования ИИ-агента. Он относится прежде всего к потоку однородных операций, где одна и та же логика агента может повторяться на большом числе заявок, платежей или проверяемых случаев. Именно в таких условиях появляется опасность серийного распространения ошибки до момента её выявления. Если процедура редкая, индивидуальная или каждый раз подтверждается специалистом, вероятность накопления повторяемой ошибки ниже, а допустимое значение α может находиться ближе к полному поручению действия агенту.

Границы применения предложенного подхода зависят от характера банковского процесса. Для справочных действий, где агент только готовит сведения и не влияет на дальнейшее проведение операции, полученный вывод имеет ограниченное прикладное значение. Для кредитования, расчётов, внутреннего контроля и спорных клиентских

ситуаций предложенная конструкция работает сильнее, поскольку ошибка там может перейти в будущие потери, дополнительные проверки или исправление ранее выполненных действий. Поэтому при практическом использовании данной схемы необходимо заранее указывать, к каким процедурам она применяется.

Отдельная трудность связана с измерением $E(\alpha)$ и $C(\alpha)$. На практике эти зависимости нельзя определить без данных о затратах на обработку, повторных рассмотрениях, жалобах, исправлениях, потерях, капитале и внутренних проверках. При отсутствии таких сведений расчёт будет опираться на экспертные допущения, что снижает правильность прикладной оценки для конкретной кредитной организации.

Значение α зависит также от качества наблюдения за агентом. Кредитная организация с быстрым выявлением отклонений может поручать агенту более широкий круг действий, поскольку период скрытого повторения ошибки будет короче. Организация с медленной реакцией должна выбирать более низкое значение α , поскольку ошибочная логика дольше остаётся внутри потока операций. Поэтому предельный уровень участия ИИ-агента определяется не только свойствами самого агента, но и скоростью внутреннего контроля.

Полученный вывод нельзя механически переносить между кредитными организациями. Один и тот же агент может иметь разное допустимое значение α из-за различий в портфеле, клиентской базе, качестве данных, внутренних правилах, скорости проверки и допустимой цене ошибки. Поэтому статья предлагает не универсальное числовое значение, а порядок рассуждения, позволяющий определить допустимый уровень участия ИИ-агента на материалах конкретной кредитной организации.

Выводы

В статье обоснован парадокс участия ИИ-агента в банковских процессах. Расширение роли агента способно снижать текущие затраты, ускорять прохождение заявок, платежей и проверок, а также уменьшать объём повторяющихся действий. Одновременно возрастает цена ошибочной логики, если она проходит через поток операций до момента выявления. Для кредитной организации это создаёт особую форму затрат, связанную с ожидаемыми потерями, капиталом, фондированием, внутренним контролем, переобучением, инцидентами и последующим исправлением.

Формальная конструкция с возрастающей вогнутой зависимостью $E(\alpha)$ и возрастающей выпуклой $C(\alpha)$ позволяет обосновать существование единственного внутреннего значения α^* , при котором совокупная оценка применения агента достигает оптимума. В этой точке предельный прирост результата совпадает с предельным ростом затрат и потерь от повторяемой ошибочной логики.

Полученный вывод показывает, что полное поручение ИИ-агенту не должно рассматриваться как естественная цель внедрения. Оправданным является не максимальное исключение работника банка из процесса, а такое значение α , при котором доход ещё покрывает ожидаемые потери, внутренние проверки и исправление ошибок.

Практическая ценность предложенной конструкции состоит в возможности определить допустимый уровень участия ИИ-агента на материалах кредитной организации. Для последующей проверки необходимо сопоставить α с долей повторных рассмотрений, числом жалоб, отклонением фактических потерь от ожидаемых, частотой инцидентов и расходами на исправление. Такая проверка позволит перейти от теоретического обоснования к эмпирической оценке применения ИИ-агента в разных банковских процессах.

Литература

1. *Diamond D. W.* Financial intermediation as delegated monitoring: a simple example // Federal Reserve Bank of Richmond Economic Quarterly. 1996. Vol. 82, No. 3. pp. 51–82. URL: https://www.richmondfed.org//media/richmondfedorg/publications/research/economic_quarterly/1996/summer/pdf/diamond.pdf.
2. *Diamond D. W., Dybvig P. H.* Bank runs, deposit insurance, and liquidity // Journal of Political Economy. 1983. Vol. 91, No. 3. pp. 401–419. DOI: 10.1086/261155
3. *Andolfatto D., Berentsen A., Martin F. M., Zhang S.* Diamond-Dybvig and beyond. On banking instability // European Economic Review. 2023. Vol. 154. Article 104437. DOI: 10.1016/j.eurocorev.2023.104437
4. *McLeMore P., Mihov A., Sanz L.* Artificial intelligence and operational losses. Evidence from U.S. bank holding companies / Federal Reserve Bank of Boston. Working paper. 2025. URL: https://www.bostonfed.org/-/media/Documents/events/2025/stress-testing-research-conference/McLeMore_Alan-dOpLosses.pdf.
5. *Ciumara T.* The impact of artificial intelligence on credit risk assessment and business model transformation in the financial sector // Annals of the «Constantin Brâncuși» University of Târgu-Jiu, Economy Series. 2024. Vol. I, Issue 6. pp. 199–203.
6. *Zhang S., Yu L.* Consumer credit risk assessment. A review from the state-of-the-art classification algorithms, data traits and learning methods // Expert Systems with Applications. 2024. Vol. 237. Article 121484. DOI 10.1016/j.eswa.2023.121484
7. *Haque Mukit M. M., Hasan F., Choudhury T., Fadli A., Fadul A.* Machine Learning & Artificial Intelligence Powered Credit Scoring Models for Islamic Microfinance Institutions: A Blockchain Approach // Risks. 2026. Vol. 14, Issue 1. DOI: 10.3390/risks14010012

8. *Awadelkarim A. M.* Federated and Interpretable AI Framework for Secure and Transparent Loan Default Prediction in Financial Institutions // *Mathematical and Computational Applications*. 2026. Vol. 31, Issue 2. DOI: 10.3390/mca31020056

9. *Bahloul R., Nabil H., Elmedany W.* Performance, Fairness, and Explainability in AI-Based Credit Scoring: A Systematic Literature Review // *Journal of Risk and Financial Management*. 2026. Vol. 19, Issue 2. DOI: 10.3390/jrfm19020104

10. *Sadok H., Sakka F., Maknoui M. E. H.* Artificial intelligence and bank credit analysis. A review // *Cogent Economics & Finance*. 2022. Vol. 10, Issue 1. Article 2023262. DOI: 10.1080/23322039.2021.2023262

11. *Danielsson J., Uthemann A.* Artificial intelligence and financial crises // *Journal of Financial Stability*. 2025. Vol. 80. Article 101453. DOI: 10.1016/j.jfs.2025.101453

12. *Bernardelli A., Adelaide Emma, Giudici P.* AI Risk Management: A Bibliometric Analysis // *Risks*. 2025. Vol. 13, Issue 7. DOI: 10.3390/risks13070131

13. *Ganesh S., Ardon L., Borrajo D., Garg D. et al.* Generative AI Agents for Knowledge Work Augmentation in Finance // *Annual Review Control, Robotics, and Autonomous Systems*. 2025. Vol. 8. PP. 189–210. DOI: 10.1146/annurev-control-022823-033813

14. *McClellan M.* AI and Financial Fragility: A Framework for Measuring Systemic Risk in Deployment of Generative AI for Stock Price Predictions // *Journal of Risk and Financial Management*. 2025. Vol. 18, Issue 9. DOI: 10.3390/jrfm18090475

15. *Mathen M., Paul A.* Toward an evolving framework for responsible AI for credit scoring in the banking industry // *Journal of Information, Communication and Ethics in Society*. 2022., Vol. 23, No. 1. PP. 148–163. DOI: 10.1108/JICES-08-2024-0122

16. *Skouloudis A., Venkatraman A.* Scratching the Surface of Responsible AI in Financial Services: A Qualitative Study on Non-Technical Challenges and the Role of Corporate Digital Responsibility // *AI* 6. 2025. Vol. 6, Issue 8. DOI: 10.3390/ai6080169

Сведения об авторе

Зверькова Татьяна Николаевна – кандидат экономических наук, доцент, доцент кафедры банковского дела и страхования; Оренбургский государственный университет, Оренбург, Россия.

ORCID: 0000-0002-6540-6154

SPIN-код: 3385-9234

E-mail: tnzverkova@mail.ru

THE LIMITS OF AI AGENT AUTONOMY IN A BANK: THE ECONOMIC PARADOX OF COST REDUCTION AND LOSS GROWTH

T. Zverkova

*Orenburg State University,
Orenburg, Russia*

Abstract. *The objective of this study is to justify the maximum level of AI agent involvement in banking operations, at which further reduction of the employee's role ceases to improve the economic evaluation of the agent's use. The methodological basis of the study includes mathematical modeling of the relationship between the degree of AI agent involvement, positive results, and the costs incurred when a repeatable error occurs. Results. It is shown that expanding the role of the AI agent reduces ongoing costs and accelerates the execution of standard banking operations, but simultaneously increases the cost of erroneous logic if it is reproduced in the flow of homogeneous operations until it is detected. It is substantiated that the benefit of delegating decisions to the agent increases with diminishing returns, while costs increase as one approaches full autonomy. Conclusions. Completely delegating operations to an AI agent should not be considered a bank's preferred option. The most economically feasible level of agent involvement is one at which the increased revenue still outweighs the increased costs associated with repeatable errors.*

Keywords: AI agent, banking operations, repeatable error, expected losses.

References

1. Diamond D. W. Financial intermediation as delegated monitoring: a simple example, *Federal Reserve Bank of Richmond Economic Quarterly*, 1996, Vol. 82,

No. 3, pp. 51-82. URL: https://www.richmondfed.org/media/richmondfedorg/publications/research/economic_quarterly/1996/summer/pdf/diamond.pdf.

2. Diamond D. W., Dybvig P. H. Bank runs, deposit insurance, and liquidity, *Journal of Political*

Economy, 1983, Vol. 91, No. 3, pp. 401-419. DOI: 10.1086/261155

3. Andolfatto D., Berentsen A., Martin F. M., Zhang S. Diamond-Dybvig and beyond. On banking instability, *European Economic Review*, 2023, Vol. 154, Article 104437. DOI: 10.1016/j.euroecorev.2023.104437

4. McLeMore P., Mihov A., Sanz L. *Artificial intelligence and operational losses. Evidence from U.S. bank holding companies* [Federal Reserve Bank of Boston. Working paper. 2025]. URL: https://www.bostonfed.org/-/media/Documents/events/2025/stress-testing-research-conference/McLeMore_AlandOpLosses.pdf.

5. Ciumara T. The impact of artificial intelligence on credit risk assessment and business model transformation in the financial sector, *Annals of the «Constantin Brâncuși» University of Târgu-Jiu, Economy Series*, 2024, Vol. I, Issue 6, pp. 199–203.

6. Zhang S., Yu L. Consumer credit risk assessment. A review from the state-of-the-art classification algorithms, data traits and learning methods, *Expert Systems with Applications*, 2024, Vol. 237, Article 121484. DOI: 10.1016/j.eswa.2023.121484

7. Haque Mukit M. M., Hasan F., Choudhury T., Fadli A., Fadul A. Machine Learning & Artificial Intelligence Powered Credit Scoring Models for Islamic Microfinance Institutions: A Blockchain Approach, *Risks*, 2026, Vol. 14, Issue 1. DOI: 10.3390/risks14010012

8. Awadelkarim A. M. Federated and Interpretable AI Framework for Secure and Transparent Loan Default Prediction in Financial Institutions, *Mathematical and Computational Applications*, 2026, Vol. 31, Issue 2. DOI: 10.3390/mca31020056

9. Bahloul R., Nabil H., Elmedany W. Performance, Fairness, and Explainability in AI-Based Credit Scoring:

A Systematic Literature Review, *Journal of Risk and Financial Management*, 2026, Vol. 19, Issue 2. DOI: 10.3390/jrfm19020104

10. Sadok H., Sakka F., Maknoui M. E. H. Artificial intelligence and bank credit analysis. A review, *Cogent Economics & Finance*. 2022. Vol. 10, Issue 1, Article 2023262. DOI: 10.1080/23322039.2021.2023262

11. Danielsson J., Uthemann A. Artificial intelligence and financial crises, *Journal of Financial Stability*, 2025, Vol. 80. Article 101453. DOI: 10.1016/j.jfs.2025.101453

12. Bernardelli A., Adelaide Emma, Giudici P. AI Risk Management: A Bibliometric Analysis, *Risks*, 2025, Vol. 13, Issue 7. DOI: 10.3390/risks13070131

13. Ganesh S., Ardon L., Borrajo D., Garg D., et al. Generative AI Agents for Knowledge Work Augmentation in Finance, *Annual Review Control, Robotics, and Autonomous Systems*, 2025, Vol. 8, pp. 189-210. DOI: 10.1146/annurev-control-022823-033813

14. McClellan M. AI and Financial Fragility: A Framework for Measuring Systemic Risk in Deployment of Generative AI for Stock Price Predictions, *Journal of Risk and Financial Management*, 2025, Vol. 18, Issue 9. DOI: 10.3390/jrfm18090475

15. Mathen M., Paul A. Toward an evolving framework for responsible AI for credit scoring in the banking industry, *Journal of Information, Communication and Ethics in Society*, 2025, Vol. 23, No. 1, pp. 148-163/ DOI: 10.1108/JICES-08-2024-0122

16. Skouloudis A., Venkatraman A. Scratching the Surface of Responsible AI in Financial Services: A Qualitative Study on Non-Technical Challenges and the Role of Corporate Digital Responsibility, *AI* 6, Vol. 6, Issue 8. DOI: 10.3390/ai6080169

About the author

Tat'yana N. Zver'kova – Candidate of Economic Sciences, Associate Professor, Associate Professor of the Department of Banking and Insurance; Orenburg State University, Orenburg, Russia.

ORCID: 0000-0002-6540-6154

SPIN-код: 3385-9234

E-mail: tnzverkova@mail.ru