Продуктовые метрики оценки эффективности Generative Ai в B2B SaaS-системах

Милосердов Андрей Андреевич
независимый исследователь, andrei2.miloserdov@gmail.com

Цель исследования состоит в развитии методического подхода к оценке эффективности Generative AI в B2B SaaS-системах на основе взаимосвязанного анализа принятия функциональности, качества взаимодействия и изменения продуктовых и бизнес-показателей. Методология включает анализ рецензируемых научных публикаций, систематизацию продуктовых метрик, структурно-логическое моделирование и сопоставление результатов. В исследовании сохранена трехуровневая модель оценки GenAI, дополненная определением объекта измерения, состава исходных данных, базового уровня, способа подтверждения эффекта и правил принятия продуктовых решений. Систематизированы показатели проникновения и устойчивости использования GenAI, практической пригодности результатов, операционной эффективности, удержания клиентов, расширения лицензий, выручки и стоимости обслуживания. Установлено, что отдельные метрики не обеспечивают достаточного основания для оценки: высокая пользовательская активность может сочетаться с интенсивным редактированием результатов, а сокращение времени выполнения задач – с недопустимым уровнем ошибок. Анализ опубликованных исследований показывает, что GenAI способен повышать производительность работников и качество отдельных видов интеллектуальной деятельности, однако величина эффекта различается по квалификации пользователей, типам задач и условиям внедрения. Предложен фреймворк, связывающий границы оценки, источники данных, три уровня метрик, проверку причинной связи и управленческие решения о масштабировании, доработке, ограничении или прекращении использования GenAI-функциональности.

Ключевые слова: Generative AI, B2B SaaS, продуктовые метрики, продуктовая аналитика, принятие функциональности, качество взаимодействия, продуктовая эффективность, бизнес-KPI, управление цифровым продуктом.

Интеграция генеративного искусственного интеллекта, или Generative AI, в B2B SaaS-системы представляет собой новый этап развития корпоративных цифровых продуктов. Сегодня GenAI применяется для подготовки документов, анализа обращений, формирования рекомендаций, поиска информации, поддержки продаж и обслуживания клиентов. Внедрение данных функций меняет ценностное предложение продукта и может влиять на способы получения дохода, распределение операций и взаимодействие с клиентами [1]. Вместе с тем наличие генеративной модели в составе программного продукта еще не подтверждает ее практическую результативность.

Вероятностная природа GenAI создает дополнительные требования к продуктовой аналитике. Сгенерированные ответы могут различаться при близких исходных условиях, содержать фактические ошибки и требовать пользовательского редактирования. Поэтому технические показатели модели следует соотносить с назначением функции, условиями ее использования и последствиями для бизнес-процесса [2].

Традиционные показатели (точность, полнота, перплексия и сходство текстов) характеризуют отдельные свойства модели; их применение не позволяет в полной мере определить, включена ли GenAI-функциональность в рабочие процессы, насколько пригодны ее результаты и влияет ли она на удержание клиентов, расширение подписки или стоимость обслуживания. Современные исследования в области качества GenAI подтверждают необходимость связывать метрики с конкретными характеристиками системы, типом результата и рисками сценария применения [3]. Соответственно, сохраняется методический разрыв между оценкой модели и оценкой продукта. Проблема приобретает особую роль в B2B SaaS, в которых пользователи имеют разные роли, полномочия и критерии результата.

Цель исследования состоит в развитии фреймворка оценки эффективности GenAI в B2B SaaS-системах на основе продуктовых метрик, отражающих принятие функциональности, качество взаимодействия и влияние на ключевые продуктовые и бизнес-показатели.

В исследовании применены анализ, синтез, сравнение, систематизация и структурно-логическое моделирование. Источниковую основу составили рецензируемые публикации по Generative AI, продуктовой аналитике, оценке качества программных систем, принятию ИИ и цифровому преобразованию организаций.

Так, оценка GenAI в корпоративном программном продукте должна охватывать последовательность формирования результата, в чем изначально определяется, используют ли функцию целевые сотрудники, далее оцениваются полезность и надежность ответа и на заключительном уровне анализируется изменение показателей продукта и клиентской организации [4]. При этом описанные уровни взаимосвязаны с точки зрения аналитики. Организационная готовность влияет на возможность результативного внедрения AI. К важным условиям относятся соответствие технологии бизнес-задаче, наличие данных, ресурсов, компетенций, поддержки руководства и понятной ответственности [5]. Поэтому продуктовые метрики следует анализировать по ролям, подразделениям, типам задач и клиентским сегментам

Первый уровень оценки характеризует фактическое включение GenAI в повседневные процессы; так, например, техническая доступность интеллектуального модуля не означает его стабильного применения, а метрики принятия позволяют разграничить первичный интерес, эпизодическое использование и регулярную работу с функцией (табл. 1).

Таблица 1
Метрики принятия функциональности GenAI в B2B SaaS-системах

Группа метрик	Метрика	Содержание	Интерпретация для B2B	Формула
Проникновение использования	Доля активных пользователей GenAI	Доля пользователей, применявших функцию в отчетном периоде	Показывает охват функциональности в клиентской организации	$A_{AI} = \frac{U_{AI}}{U_{active}} \times 100\%$
Проникновение по ролям	Доля активных пользователей по роли	Использование GenAI среди сотрудников определенной роли	Выявляет группы, в которых формируется основная ценность	$A_r = \frac{U_{AI,r}}{U_r} \times 100\%$
Интенсивность	Частота взаимодействий	Среднее число обращений на одного пользователя	Характеризует регулярность использования	$F_{AI} = \frac{N_{\text{int}}}{U_{AI}}$

Встраивание в процесс	Доля задач с поддержкой GenAI	Часть задач определенного типа, выполненных с GenAI	Показывает место функции в рабочем процессе	$S_{task} = \frac{N_{AItask}}{N_{task}} \times 100\%$
Повторное использование	Доля повторных пользователей	Доля пользователей, повторно применявших функцию	Отражает устойчивость воспринимаемой ценности	$R_{use} = \frac{U_{repeat}}{U_{AI}} \times 100\%$
Качество принятия	Уровень отклонения результатов	Доля результатов, удаленных или полностью переписанных	Высокое значение указывает на низкую применимость или доверие	$R_{rej} = \frac{N_{rejected}}{N_{generated}} \times 100\%$
Удержание	Удержание пользователей функции	Доля пользователей, продолжающих применение через N периодов	Показывает долгосрочную интеграцию	$\mathrm{Re}\,t_{AI} = \frac{U_{t+N}}{U_t} \times 100\%$
Распределение	Доля взаимодействий по подразделениям	Распределение обращений между бизнес-подразделениями	Выявляет центры создания ценности и барьеры внедрения	$D_d = \frac{N_{\mathrm{int,d}}}{N_{\mathrm{int}}} \times 100\%$

Представленные в таблице 1 показатели позволяют определить, насколько GenAI-функциональность включена в рабочие процессы пользователей. Интерпретация метрик предполагает совместное рассмотрение, поскольку высокое значение одной метрики еще не подтверждает результативность внедрения. Большая доля активных пользователей при низком уровне повторного использования может отражать первоначальный интерес к новой функции, который не перешел в регулярное применение, а высокая частота обращений при существенной доле отклоненных результатов, в свою очередь, указывает на необходимость повторных генераций и дополнительные затраты времени.

Более убедительным свидетельством принятия становится сочетание нескольких признаков, а именно удержания пользователей функции, роста доли задач с поддержкой GenAI и распространения технологии среди различных профессиональных ролей. Подобное сочетание показывает, что функция используется для решения повторяющихся задач и постепенно становится частью установленного рабочего процесса. При анализе необходимо учитывать различия между подразделениями, должностями и группами пользователей, поскольку одинаковая функциональность может создавать неодинаковую ценность для операционных сотрудников, руководителей и специалистов, принимающих решения.

Значение дифференцированной оценки подтверждается результатами исследования внедрения GenAI-помощника в деятельность 5172 специалистов клиентской поддержки. Использование технологии повысило производительность работников в среднем приблизительно на 15%, причем наиболее высокий результат был зафиксирован среди менее опытных сотрудников. Эффект у квалифицированных специалистов оказался слабее, что указывает на зависимость результата от исходного уровня знаний и характера выполняемых задач [6]. Следовательно, общая доля активных пользователей должна дополняться показателями принятия по ролям, стажу, квалификации и типам операций.

Метрики первого уровня позволяют установить наличие использования, однако не формируют полной характеристики эффективности. Пользователь может регулярно обращаться к GenAI и при этом тратить значительное время на проверку и редактирование ответа. Поэтому после оценки принятия необходимо определить качество взаимодействия и практическую пригодность получаемых результатов (таблица 2):

Таблица 2
Метрики качества взаимодействия и функциональной полезности GenAI

Группа метрик	Метрика	Содержание	Интерпретация для B2B	Формула
Практическая применимость	Уровень принятия результата	Доля сгенерированных результатов, использованных без существенной переработки	Показывает непосредственную пригодность ответа для бизнес-процесса	$AR = \frac{N_{accepted}}{N_{generated}} \times 100\%$
Практическая применимость	Степень редактирования	Доля элементов результата, измененных пользователем перед применением	Характеризует объем дополнительной работы после генерации	$ED = \frac{N_{edited}}{N_{elements}} \times 100\%$
Надежность	Уровень критических ошибок	Доля проверенных результатов с существенными фактическими, смысловыми или регуляторными ошибками	Определяет допустимость применения функции в чувствительных процессах	$ER = \frac{N_{critical}}{N_{checked}} \times 100\%$
Операционная результативность	Сокращение времени выполнения задачи	Изменение продолжительности работы относительно базового уровня	Отражает влияние функции на производительность пользователя	$TS = \frac{T_0 - T_{AI}}{T_0} \times 100\%$
Трудоемкость взаимодействия	Среднее число итераций	Количество запросов, повторных генераций и уточнений до получения приемлемого результата	Показывает дополнительные усилия пользователя	$I = \frac{N_{iterations}}{N_{tasks}}$
Качество первого результата	Пригодность первого ответа	Доля задач, завершенных без повторной генерации	Характеризует точность первоначального соответствия запросу	$FUR = \frac{N_{first}}{N_{generated}} \times 100\%$
Пользовательская оценка	Удовлетворенность результатом	Среднее значение оценки сгенерированных ответов пользователями	Отражает воспринимаемое качество взаимодействия	$CSAT = \frac{\sum_{j=1}^{n} s_j}{n}$
Пользовательская оценка	Воспринимаемая полезность	Средняя оценка влияния GenAI на выполнение профессиональной задачи	Показывает готовность пользователя продолжать применение функции	$PU = \frac{\sum_{j=1}^{n} p_j}{n}$

Показатели таблицы 2 образуют связанную систему. Высокий уровень принятия результата приобретает положительное значение при низкой степени редактирования и небольшом числе повторных итераций; соответственно, в таком сочетании GenAI сокращает объем ручной работы и ускоряет завершение задачи. Если пользователи формально принимают ответы, но вынуждены вносить многочисленные изменения, фактический эффект оказывается ниже значения, которое показывает частота использования.

Особое внимание необходимо уделять критическим ошибкам. Сокращение времени выполнения не может рассматриваться как достаточный результат, если использование GenAI повышает риск неверных решений, нарушений требований или распространения недостоверных сведений. Для каждого бизнес-процесса должен быть установлен допустимый уровень ошибок, превышение которого служит основанием для ограничения функции, усиления проверки либо изменения сценария ее применения.

Необходимость совместного измерения времени и качества подтверждается экспериментом с участием 453 специалистов, выполнявших профессиональные письменные задания. Применение генеративного ИИ позволило сократить продолжительность работы примерно на 40% и одновременно повысить оценку качества результата на 18% [7]. Соответственно, оценка производительности должна учитывать итоговую пригодность выполненной работы, поскольку ускорение при снижении качества не создает устойчивой ценности для корпоративного продукта.

Содержание задачи также влияет на выбор показателей. В исследовании применения GenAI в B2B-продажах установлена положительная связь использования технологии с результативностью процесса продаж, административной эффективностью и показателями деятельности сотрудников. При этом выраженность результата зависела от поддержки руководства и уверенности работников в использовании технологии [8]. Следовательно, метрики качества взаимодействия целесообразно анализировать вместе с характеристиками организационной среды и подготовленности пользователей.

Распределение операций между сотрудником и интеллектуальной системой должно учитывать характер требуемой деятельности, а именно: повторяющиеся аналитические задачи и обработка структурированной информации обладают более высоким потенциалом автоматизации; процессы, связанные с эмоциональным взаимодействием, неоднозначными решениями и повышенной ответственностью, зависят от более выраженного участия специалиста [9]. В подобных сценариях GenAI может выполнять вспомогательную функцию, но окончательное решение сохраняется за пользователем. Так, исследования клиентского опыта показывают, что оценка AI-взаимодействия зависит от качества сервиса, доверия, понятности результата и способности технологии облегчить достижение цели пользователя [10]. Намерение продолжать использование интеллектуальной функции формируется под влиянием ожидаемой результативности, удобства, эмоциональной реакции и воспринимаемых рисков [11]. По этой причине данные опросов следует сопоставлять с фактическим поведением пользователей,

поскольку высокая субъективная оценка не всегда сопровождается регулярным применением.

После определения качества взаимодействия возникает возможность перейти к третьему уровню оценки (таблица 3), который характеризует влияние GenAI-функциональности на состояние клиентской базы, масштабы использования SaaS-системы, выручку и операционные расходы. Показатели данного уровня обладают большей управленческой значимостью, однако требуют осторожной интерпретации.

Таблица 3
Продуктовые и бизнес-KPI оценки GenAI в B2B SaaS-системах

Группа метрик	Метрика	Содержание	Интерпретация для B2B	Формула
Удержание клиентов	Уровень удержания	Доля клиентов, продолжающих использовать продукт в конце периода	Показывает устойчивость создаваемой продуктом ценности	$RR = \frac{C_{retained}}{C_0} \times 100\%$
Удержание клиентов	Уровень оттока	Доля клиентов, прекративших использование продукта	Характеризует стабильность клиентской базы	$ChR = \frac{C_{lost}}{C_0} \times 100\%$
Расширение использования	Рост числа лицензий	Изменение количества активных лицензий в клиентской организации	Отражает внутреннее распространение продукта и GenAI-функции	$LE = \frac{L_1 - L_0}{L_0} \times 100\%$
Динамика выручки	Чистое удержание выручки	Изменение регулярной выручки с учетом расширения, сокращения и оттока	Позволяет оценить изменение ценности клиентского портфеля	$NRR = \frac{MRR_0 + E - C - Ch}{MRR_0} \times 100\%$
Динамика выручки	Рост ARPU	Изменение средней выручки на одного пользователя	Показывает возможность монетизации GenAI-функции	$ARPU_g = \frac{ARPU_1 - ARPU_0}{ARPU_0} \times 100\%$
Операционная результативность	Сокращение времени выполнения	Изменение продолжительности клиентского процесса после внедрения GenAI	Отражает результат для организации-пользователя	$TS = \frac{T_0 - T_{AI}}{T_0} \times 100\%$
Качество процесса	Сокращение ошибок	Изменение частоты ошибок в операциях с поддержкой GenAI	Показывает влияние функции на надежность процесса	$ERed = \frac{ER_0 - ER_{AI}}{ER_0} \times 100\%$
Стоимость обслуживания	Снижение cost-to-serve	Изменение расходов поставщика на обслуживание одного клиента	Характеризует влияние функции на экономику SaaS-продукта	$CSR = \frac{CTS_0 - CTS_{AI}}{CTS_0} \times 100\%$
Финансовый результат	Возврат инвестиций в GenAI	Соотношение финансового результата и затрат на разработку и эксплуатацию функции	Используется при принятии решения о дальнейшем развитии	$ROI_{AI} = \frac{\Delta GP + S - C_{AI}}{C_{AI}} \times 100\%$

Приведенные метрики следует сопоставлять с данными двух предыдущих уровней. Так, рост NRR или количества лицензий имеет более выраженную связь с GenAI, если функция регулярно применяется целевыми пользователями, а ее результаты обладают подтвержденной полезностью. При отсутствии данной связи изменение финансовых показателей может отражать общую динамику продукта. Влияние ИИ на результаты организации формируется при объединении технологических ресурсов, данных, профессиональных компетенций и управленческих возможностей. Наличие технологии не обеспечивает улучшения результатов деятельности, т.к. свою роль играют умение внедрять ее в процессы, качество данных и способность сотрудников применять получаемые результаты. Данное положение подтверждает необходимость рассматривать продуктовые и бизнес-KPI в связи с условиями использования функции.

Для установления вклада GenAI следует применять методы сопоставления. Наиболее надежные результаты обеспечивает A/B-тестирование, при котором пользователи или клиентские организации распределяются между группами с различным доступом к функции. Если проведение контролируемого эксперимента невозможно, допустимо сравнение сопоставимых когорт, анализ периодов до и после внедрения, а также учет сезонности, состава клиентов и изменений тарифа.

Цифровое преобразование продукта затрагивает его ценностное предложение, способы взаимодействия с клиентами и механизмы формирования выручки. Поэтому изменение одного показателя редко может быть объяснено единственным фактором; оценка должна опираться на последовательную проверку того, каким образом использование GenAI изменило действия пользователя, результат задачи и итоговый KPI. Важно учитывать и положение функции внутри общего бизнес-процесса, т.к. локальное ускорение отдельной операции может не привести к сокращению продолжительности всего процесса, если последующие стадии сохраняют прежние ограничения. Ценность технологии формируется через изменения в операциях, организационном взаимодействии и способах создания результата. Следовательно, продуктовые команды должны оценивать GenAI в пределах завершенного пользовательского сценария, а не отдельного обращения к модели.

Совместное рассмотрение принятия, качества и бизнес-эффекта позволяет избежать односторонней оценки. Высокое использование при слабой экономической отдаче указывает на необходимость снижения стоимости, изменения тарифа или выбора более значимого сценария. Положительный эффект среди ограниченной группы пользователей при низком общем принятии свидетельствует о барьерах освоения. Низкие показатели всех трех уровней служат основанием для существенной переработки функции.

Исходя из этого, нами предлагается фреймворк интегрированной оценки GenAI в B2B SaaS (рис. 1):

На первом этапе требуется точно определить объект оценки. Использование общего обозначения «GenAI в продукте» затрудняет интерпретацию, поскольку одна SaaS-система может включать несколько функций с различными задачами и уровнями риска. Поэтому отдельно оцениваются подготовка текста, поиск информации, классификация обращений, формирование рекомендации или иной завершенный сценарий. Фиксация базового уровня позволяет установить исходные значения времени, ошибок, пользовательской активности и бизнес-KPI. Без данного сопоставления невозможно определить, произошло ли улучшение после внедрения функции. Следующие три этапа соответствуют уровням принятия, качества и продуктово-экономического эффекта.

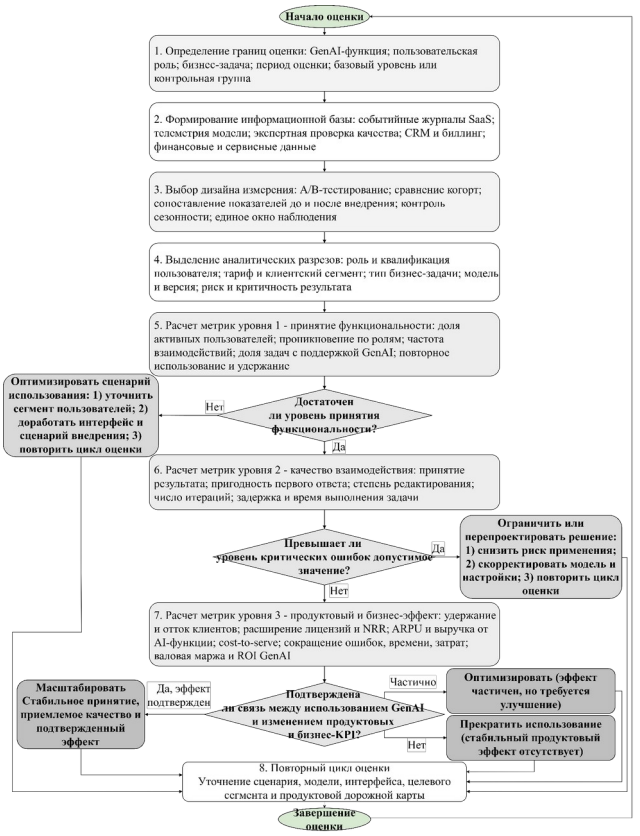


Рисунок 1 – Алгоритм интегрированной оценки эффективности GenAI в B2B SaaS-системах

Межуровневая интерпретация используется для выявления причин результата. Возможны четыре основных сочетания:

Высокое принятие, приемлемое качество и подтвержденный бизнес-эффект. Такое сочетание служит основанием для масштабирования функции.

Высокое принятие при недостаточном качестве или слабом экономическом результате. Требуются оптимизация модели, сокращение количества итераций, изменение стоимости или уточнение пользовательского сценария.

Низкое принятие при положительном результате среди фактических пользователей. Основное внимание следует уделить обучению, интерфейсу и включению функции в установленный рабочий процесс.

Низкое принятие, неудовлетворительное качество и отсутствие бизнес-эффекта. Функция нуждается в существенной переработке либо прекращении использования.

Дополнительным ограничением выступает уровень критических ошибок:

$$ER \le E\,R_{max},$$

где $E\,R_{max}$ – предельно допустимая доля критических ошибок для конкретного процесса.

Если установленное условие не выполняется, функция не должна масштабироваться независимо от частоты использования и предполагаемой экономии. В данном случае приоритет получает снижение риска и повышение надежности результата.

Предлагаемый фреймворк имеет циклический характер. После изменения модели, интерфейса, тарифа или целевого сценария показатели рассчитываются повторно, что сответствует характеру цифровых преобразований, которые затрагивают организационные процессы и распределение компетенций и развиваются на протяжении продолжительного периода.

В результате трехуровневая модель превращается в последовательность управленческой оценки и позволяет определить, применяется ли функция, создает ли она пригодный результат и приводит ли ее использование к изменению показателей продукта. Полученная информация формирует основание для решения о дальнейшем развитии GenAI в B2B SaaS-системе.

Таким образом, проведенное исследование позволило развить методический подход к оценке эффективности Generative AI в B2B SaaS-системах на основе последовательного анализа принятия функциональности, качества взаимодействия и изменения продуктовых и бизнес-показателей. Установлено, что отдельные показатели пользовательской активности не позволяют сделать обоснованный вывод о результативности GenAI-функции. Высокая частота использования может сопровождаться значительным объемом редактирования, повторными генерациями и сохранением критических ошибок. По этой причине метрики принятия должны сопоставляться с показателями практической пригодности результата, трудоемкости взаимодействия и надежности. Практическое применение подхода позволяет разграничивать ситуации, требующие масштабирования функциональности, совершенствования пользовательского сценария, ограничения применения либо прекращения дальнейшего развития функции. Перспективным направлением исследования является апробация предложенного фреймворка на данных действующей B2B SaaS-системы с использованием A/B-тестирования или когортного анализа.

Литература

1. Kanbach D. K., Heiduk L., Blueher G., Schreiter M., Lahmann A. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective // Review of Managerial Science. – 2024. – Vol. 18, No. 4. – P. 1189–1220. – DOI: 10.1007/s11846-023-00696-z.
2. Feuerriegel S., Hartmann J., Janiesch C., Zschech P. Generative AI // Business & Information Systems Engineering. – 2024. – Vol. 66, No. 1. – P. 111–126. – DOI: 10.1007/s12599-023-00834-7.
3. Yu L., Alégroth E., Chatzipetrou P., Gorschek T. Measuring the quality of generative AI systems: mapping metrics to quality characteristics – snowballing literature review // Information and Software Technology. – 2025. – Vol. 186. – Art. 107802. – DOI: 10.1016/j.infsof.2025.107802.
4. Bandi A., Adapa P. V. S. R., Kuchi Y. E. V. P. K. The power of generative AI: a review of requirements, models, input–output formats, evaluation metrics, and challenges // Future Internet. – 2023. – Vol. 15, No. 8. – Art. 260. – DOI: 10.3390/fi15080260.
5. Jöhnk J., Weißert M., Wyrtki K. Ready or not, AI comes: an interview study of organizational AI readiness factors // Business & Information Systems Engineering. – 2021. – Vol. 63. – P. 5–20. – DOI: 10.1007/s12599-020-00676-7.
6. Brynjolfsson E., Li D., Raymond L. Generative AI at work // The Quarterly Journal of Economics. – 2025. – Vol. 140, No. 2. – P. 889–942. – DOI: 10.1093/qje/qjae044.
7. Noy S., Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence // Science. – 2023. – Vol. 381, No. 6654. – P. 187–192. – DOI: 10.1126/science.adh2586.
8. Rodriguez M., Deeter-Schmelz D. R., Krush M. T. The impact of generative AI technology on B2B sales process and performance: an empirical study // Journal of Business & Industrial Marketing. – 2025. – Vol. 40, No. 10. – P. 2013–2027. – DOI: 10.1108/JBIM-02-2025-0097.
9. Huang M.-H., Rust R. T. Artificial intelligence in service // Journal of Service Research. – 2018. – Vol. 21, No. 2. – P. 155–172. – DOI: 10.1177/1094670517752459.
10. Ameen N., Tarhini A., Reppel A., Anand A. Customer experiences in the age of artificial intelligence // Computers in Human Behavior. – 2021. – Vol. 114. – Art. 106548. – DOI: 10.1016/j.chb.2020.106548.
11. Gursoy D., Chi O. H., Lu L., Nunkoo R. Consumers' acceptance of artificially intelligent device use in service delivery // International Journal of Information Management. – 2019. – Vol. 49. – P. 157–169. – DOI: 10.1016/j.ijinfomgt.2019.03.008.

Product metrics for evaluating the effectiveness of generative Ai in B2B SaaS systems

Miloserdov A.I.
Independent Researcher

JEL classification: C01, C02, C1, C4, C5, C6, C8

The study aims to develop a methodological framework for evaluating the effectiveness of Generative AI in B2B SaaS systems through the interrelated analysis of feature adoption, interaction quality, and changes in product and business indicators. The methodology includes an analysis of peer-reviewed publications, systematization of product metrics, structural modeling, and comparison of published empirical results. The original three-level evaluation model is retained and extended by defining the unit of analysis, data sources, baseline indicators, methods for effect attribution, and product decision rules. The study systematizes metrics of GenAI penetration and sustained use, output applicability, operational performance, customer retention, license expansion, revenue, and cost-to-serve. The findings show that individual metrics do not provide a sufficient basis for evaluation: high user activity may be accompanied by extensive output editing, while faster task completion may coincide with an unacceptable error rate. Published empirical research indicates that GenAI can improve employee productivity and the quality of selected knowledge-work tasks. However, the magnitude of the effect varies across user skill levels, task types, and implementation conditions. The proposed framework links evaluation boundaries, data sources, three metric levels, effect attribution, and product decisions concerning scaling, improvement, restriction, or discontinuation of GenAI features.

Keywords: Generative AI, B2B SaaS, product metrics, product analytics, feature adoption, interaction quality, product performance, business KPI, digital product management.

References

1. Kanbach D. K., Heiduk L., Blueher G., Schreiter M., Lahmann A. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective // Review of Managerial Science. – 2024. – Vol. 18, No. 4. – P. 1189–1220. – DOI: 10.1007/s11846-023-00696-z.
2. Feuerriegel S., Hartmann J., Janiesch C., Zschech P. Generative AI // Business & Information Systems Engineering. – 2024. – Vol. 66, No. 1. – P. 111–126. – DOI: 10.1007/s12599-023-00834-7.
3. Yu L., Alégroth E., Chatzipetrou P., Gorschek T. Measuring the quality of generative AI systems: mapping metrics to quality characteristics – snowballing literature review // Information and Software Technology. – 2025. – Vol. 186. – Art. 107802. – DOI: 10.1016/j.infsof.2025.107802.
4. Bandi A., Adapa P. V. S. R., Kuchi Y. E. V. P. K. The power of generative AI: a review of requirements, models, input–output formats, evaluation metrics, and challenges // Future Internet. – 2023. – Vol. 15, No. 8. – Art. 260. – DOI: 10.3390/fi15080260.
5. Jöhnk J., Weißert M., Wyrtki K. Ready or not, AI comes: an interview study of organizational AI readiness factors // Business & Information Systems Engineering. – 2021. – Vol. 63. – P. 5–20. – DOI: 10.1007/s12599-020-00676-7.
6. Brynjolfsson E., Li D., Raymond L. Generative AI at work // The Quarterly Journal of Economics. – 2025. – Vol. 140, No. 2. – P. 889–942. – DOI: 10.1093/qje/qjae044.
7. Noy S., Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence // Science. – 2023. – Vol. 381, No. 6654. – P. 187–192. – DOI: 10.1126/science.adh2586.
8. Rodriguez M., Deeter-Schmelz D. R., Krush M. T. The impact of generative AI technology on B2B sales process and performance: an empirical study // Journal of Business & Industrial Marketing. – 2025. – Vol. 40, No. 10. – P. 2013–2027. – DOI: 10.1108/JBIM-02-2025-0097.
9. Huang M.-H., Rust R. T. Artificial intelligence in service // Journal of Service Research. – 2018. – Vol. 21, No. 2. – P. 155–172. – DOI: 10.1177/1094670517752459.
10. Ameen N., Tarhini A., Reppel A., Anand A. Customer experiences in the age of artificial intelligence // Computers in Human Behavior. – 2021. – Vol. 114. – Art. 106548. – DOI: 10.1016/j.chb.2020.106548.
11. Gursoy D., Chi O. H., Lu L., Nunkoo R. Consumers' acceptance of artificially intelligent device use in service delivery // International Journal of Information Management. – 2019. – Vol. 49. – P. 157–169. – DOI: 10.1016/j.ijinfomgt.2019.03.008.