

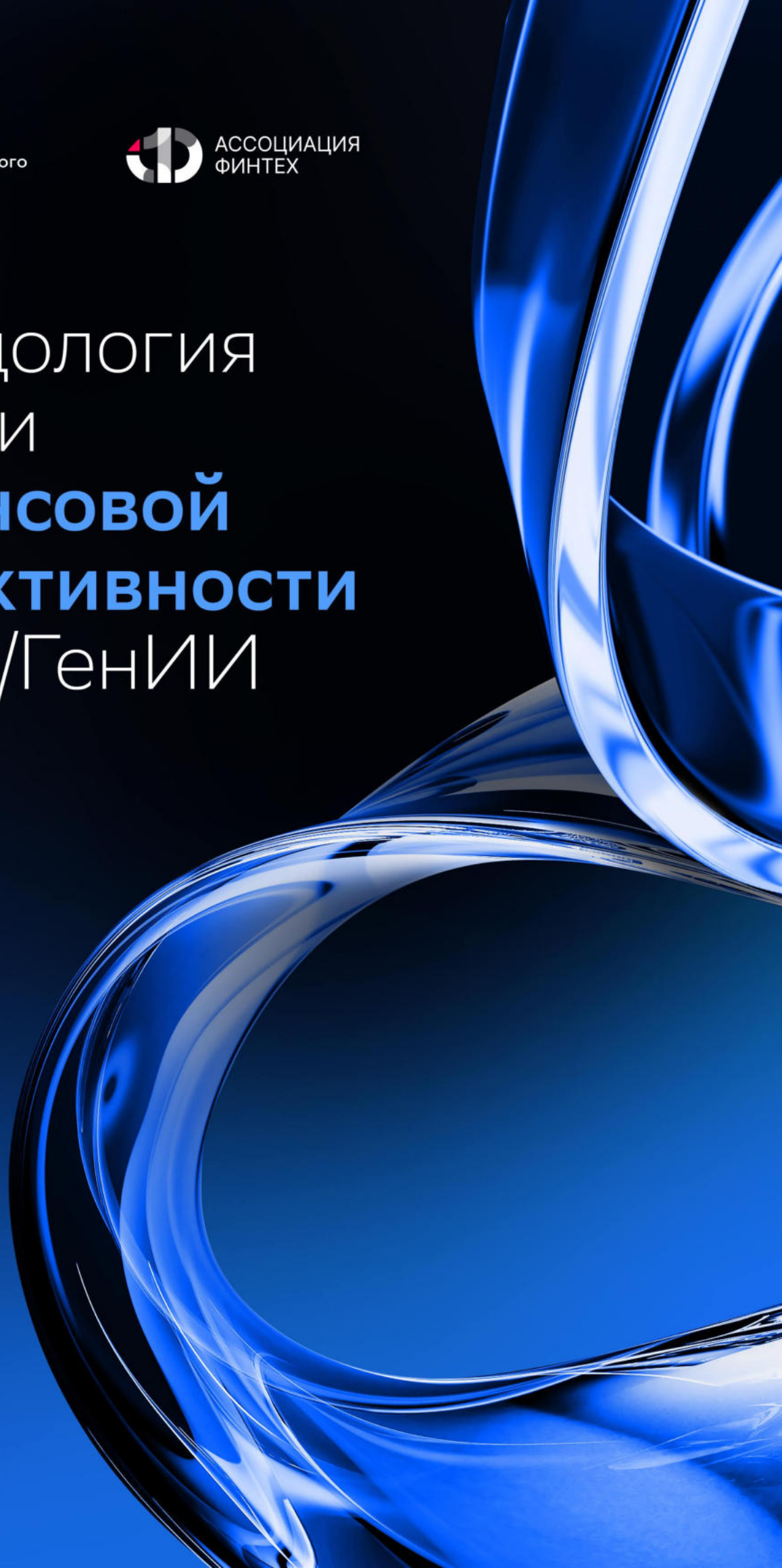


АЛЬЯНС
В СФЕРЕ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА



АССОЦИАЦИЯ
ФИНТЕХ

Методология оценки **финансовой эффективности** от ИИ/ГенИИ



При поддержке:

- ПАО «Сбербанк»
- АО «ТБАНК»
- АО «АЛЬФА-БАНК»
- ПАО «Банк ВТБ»
- ПАО Московская Биржа
- АО «Газпромбанк»
- ПАО «Промсвязьбанк»
- АО «Россельхозбанк»
- «Страховое акционерное общество «ВСК»



Валерия Воробьева,
Генеральный директор
Альянса в сфере ИИ

“

Разработанная методология оценки финансовой эффективности ИИ - это прорыв для финансовой отрасли: она выводит внедрение искусственного интеллекта из стадии разрозненных пилотов в режим системного, ответственного управления эффектом. Это не просто инструмент расчёта, а основа прозрачности, зрелости рынка и доверия. В ней собраны лучшие практики - от распределения ролей и задач до выбора метрик и методов, проверенных на реальных кейсах. Применяя эту методологию, участники отрасли подтверждают свою приверженность высоким стандартам: каждый расчёт становится частью общей доверенной среды, где за достоверностью данных и честностью оценки стоит сама методология. Именно такой подход создаёт условия для устойчивого роста инвестиций в ИИ и масштабирования его реального эффекта.

”



Максим Григорьев,
Генеральный директор
Ассоциации ФинТех

“

Финансовая отрасль – лидер с точки зрения применения технологий искусственного интеллекта. При этом бизнес сегодня внимательно относится к вопросу эффективности инвестиций и обоснованию экономического эффекта. Методология, разработанная и апробированная в рамках Отраслевого клуба, созданного при участии Ассоциации ФинТех и Альянса в сфере ИИ, унифицирует подходы по оценке финансовых эффектов применения искусственного интеллекта на основе опыта лидеров отрасли. Ее использование упростит бизнес-обоснование использования ИИ, позволит заложить основу для формирования отраслевых и кросс-отраслевых бенчмарков, будет способствовать развитию рынка ИИ-решений для финансовой отрасли.

”

Содержание

01

Что такое ИИ

4

02

Зачем и кто оценивает
фин. эффект в рамках
ИИ трансформации?

5

03

Основные принципы
оценки фин.
эффективности от ИИ

- 7 Принципы финансовой оценки
- 10 Важность изучения процессов AS IS (до внедрения ИИ) и TO BE (после внедрения ИИ)
- 12 Типы финансовых эффектов
- 25 Что такое инкремент?
- 31 Чистый (net) vs Валовый (gross)

04

Какими метриками
можно оценить
фин. эффект от ИИ?

- 33 Набор метрик для оценки
ФЭ в зависимости от целей,
задач
- 35 Как оценить собственные
цели и результаты
относительно лучших
рыночных практик?

07

Как избежать двойного
подсчета эффекта
от проектов

58

05

Какими методами
можно посчитать
эффект от ИИ?

- 39 Контролируемые
рандомизированные
эксперименты
- 49 Квази-экспериментальные
методы
- 53 Альтернативное
моделирование

08

ТОП-ошибок
при расчете эффекта

65

06

Предлагаемые решения
сложных вопросов на
базе best practices
лидеров индустрии

- 54 Как измерять эффект, где
нет денежных метрик?
Косвенные стратегические
метрики
- 55 Как определить тип и
объем инвестиций в R&D?

09

Кейсы с примерами
оценки фин. эффектов
от отдельных внедрений

69

Активное развитие и внедрение технологий искусственного интеллекта (далее - ИИ) оказывают существенное влияние на финансовый рынок. ИИ-технологии применяются в кредитном скоринге, управлении рисками, инвестиционных решениях, противодействии мошенничеству и других ключевых процессах. Вместе с новыми возможностями они несут и серьезные этические риски.

В ответ на эти вызовы Банком России по результатам консультаций с участниками финансового рынка, в том числе с отраслевым клубом «ИИ в финансовой отрасли», созданным Альянсом в сфере ИИ и Ассоциацией ФинТех, был подготовлен Кодекс этики в сфере разработки и применения искусственного интеллекта на финансовом рынке¹.

Настоящая Методология оценки финансовой эффективности от ИИ и Генеративного ИИ (далее - ГенИИ) разработана в логике положений Кодекса. Она призвана дополнить этический фундамент практическим инструментарием, обеспечивающим объективность и воспроизводимость оценки экономического эффекта ИИ-инициатив. Такой подход позволяет организациям финансового рынка учитывать не только коммерческую отдачу проектов, но и согласовывать внедрение ИИ-технологий с принципами ответственного и добросовестного применения.

¹ Информационное письмо Банка России от 09.07.2025 № ИН-016-13/91

01 / Что такое ИИ?

ИИ/ML (технологии машинного обучения) – это алгоритмы на базе обучающейся математической модели, которые позволяют выполнять задачи, требующие человеческого интеллекта, на основе анализа данных

Эффект от ИИ/ML возникает только при внедрении модели в бизнес-процесс для решений поставленных задач

Примеры:

Виртуальные ассистенты на базе ИИ

Отвечают на вопросы клиентов, помогают с банковскими операциями и предоставляют необходимую информацию

Антифрод и финансовый мониторинг

ИИ анализирует нетипичное поведение физических и юридических лиц, чтобы противостоять финансовому мошенничеству

Кредитный скоринг

ИИ анализирует данные о заёмщике, чтобы оценить кредитоспособность и предсказать риск невозврата средств и т.д.

Зачем и кто оценивает фин. эффект в рамках ИИ трансформации?

Цель

Внутри компании:

- Обоснование портфеля ИИ-инициатив при помощи:
 - Оценки ROI ИИ-инициатив
 - Приоритизации ИИ-инициатив - Классического ML и ГенИИ²
- Проверка соответствия инициатив стратегическим целям
- Приоритизация ресурса для внедрения ИИ³
- Управление бизнес-рисками⁴

Роли и ответственные:

	Бизнес	Фин. валидация
«Новички» ⁵		
«Средние» ⁵		
«Продвинутые» ⁵		

 - Обязательные роли

 - Опциональные роли

Для внешнего рынка:

- Усиление репутации
- Увеличение стоимости бренда
- Привлечение / удержание акционеров

Тех. валидация

AI - стратегия

Аудит



2 Н-р, для Классического ML – на базе ROI, прибыли до налога ближайшего года / 3х лет, для ГЕНИИ (Генеративного ИИ) – на базе более долгосрочных метрик, например, CLTV

3 Инфраструктура & люди

4 Н-р, продажи, производительность,...

5 Уровень зрелости компании определяется внутри экспертно, методология оценки - к дальнейшей проработке

Основные принципы оценки финансовой эффективности от ИИ

Принципы финансовой оценки

Важно понимать, что финансовая оценка – это не набор универсальных математических законов и операций. Финансовая оценка – это обоснованное мнение, подкрепленное расчетами. По определению, мнение подразумевает субъективность, которую необходимо минимизировать для принятия взвешенных управленческих решений.

Для этой цели разработан набор рекомендаций, принципов финансовой оценки:

1. Принцип максимальной объективности:

В процессе финансовой оценки необходимо избегать экспертных оценок, не подкрепленных расчетами. Важно опираться на объективные данные и статистику. Иными словами, кто угодно, обладающий той же статистикой, на которой была выполнена оценка, должен иметь возможность повторить расчет.

2. Принцип измеримости:

Выгоды и затраты по проекту должны быть измеримы. Все параметры расчета должны быть зафиксированы в качестве драйверов для последующего мониторинга эффекта и проверки корректности сделанных выводов и допущений. Сюда входят как финансовые драйверы (операционный доход, ОПЕХ, свободный денежный поток и т.д.), так и нефинансовые (количество продаж, отклик, текучесть персонала, и т.д.). Иными словами, в любой момент времени жизни ИИ-инициативы, необходимо иметь возможность получить значение каждого драйвера оценки.

3. Принцип консерватизма:

Денежные потоки обладают свойством неопределенности. Необходимо выбрать единый подход к работе с неопределенностью, чтобы финансовые оценки возможно было сравнивать между собой. Чтобы минимизировать риски реализации негативных сценариев, и тем самым повысить доверие к оценке, принцип консерватизма предлагает заранее, на этапе оценки, для всех ИИ-инициатив заложить максимально негативный сценарий развития событий. Иными словами, доходы необходимо оценивать на минимально возможном уровне, а расходы – на максимально возможном.

4. Принцип связи с инициативой

В качестве выгод и затрат рассматриваются инкрементальные (дополнительные) потоки, связанные с конкретным объектом оценки. Если инициатива является не единственным фактором, влияющим на бизнес-процесс (например, изменение текущего

процесса (AS IS) при одновременной реализации инициативы; запуск множества инициатив одновременно), необходимо выделять и оценивать эффект от каждого изменения отдельно. Таким образом, можно будет корректно оценить целесообразность, а также корректно определить расходы, понесенные на реализацию конкретных решений. Если же эффект от изменения какого-либо процесса не может быть разделен между инициативами, то данная группа инициатив оценивается совместно как единый кейс (каскад моделей).

5. Принцип полноты:

Требуется учитывать все инкрементальные расходы и доходы, необходимые для достижения бизнес-цели/внедрения модели в бизнес-процесс: как по целевому продукту, так и по сплиту продаж в целом (каннибализация и кросс-селл); как в целевом канале сбыта, так и во всех доступных каналах; как прямые расходы на внедрение, так и косвенные расходы всех связанных подразделений; как текущие расходы, так и расходы на поддержку модели в будущем.

6. Принцип материальности

Нет необходимости стремиться рассчитать эффект с точностью до копейки. Необходимо ответить на вопрос, насколько чувствительна прибыль к изменению тех положительных или отрицательных драйверов, точность которых мы хотим повысить. Так, если мы видим, что значение оценки может измениться не существенно (например, на 1%, на уровне погрешности), при этом нет влияния на итоговое решение об экономической целесообразности ИИ инициативы, то использование дополнительных ресурсов для доработки системы оценки для такой точности не нужно.

7. Принцип учёта консолидированного эффекта:

Необходимо проводить оценку на уровне всей группы компаний/холдинга (если применимо), чтобы учесть все связанные финансовые и налоговые последствия.

Также основным принципом оценки финансового эффекта служит четкий и последовательный процесс оценки ожидаемого эффекта для успешного внедрения искусственного интеллекта и получения от него максимальной отдачи. Такой подход позволяет не только сформировать реалистичные прогнозы, но и вовремя скорректировать ожидания на основе реальных данных и изменений внешней среды. Ниже описаны ключевые этапы, обеспечивающие комплексную оценку и управление эффектом от ИИ.

Ключевые этапы расчета эффекта от ИИ:

1. Формулировка бизнес-гипотез

Оценка эффекта может начинаться с уровня формирования гипотезы. Изначальный прогноз вашего эффекта (гипотеза) может отличаться от фактического результата, поэтому риски таких отклонений следует закладывать с использованием принципа консервативной оценки - учитывать в прогнозе минимальные доходы и максимальные расходы.

2. Первичная оценка финансового эффекта

На данном этапе проводится начальная проверка предположений с целью оценки потенциального влияния ИИ на бизнес-процессы и выработки более обоснованных прогнозов.

3. Проведение Пилота

Чем дальше мы продвигаемся от идеи к практике, тем точнее становится оценка, так как появляются пилотные данные. Пилот позволяет проверить рабочие гипотезы в реальных условиях, с возможностью учета внешних изменений, например, изменения ключевой ставки ЦБ, которые могут повлиять на результаты.

4. Оценка эффекта

После завершения пилотного этапа производится более точный анализ полученных результатов, позволяющий оценить реальный вклад ИИ в бизнес с учетом всех выявленных факторов.

5. Мониторинг эффекта

На этом этапе осуществляется постоянный контроль и анализ эффективности внедренного решения, чтобы своевременно реагировать на возможные изменения внешних и внутренних условий и поддерживать высокую точность оценки.

Таким образом, по мере прохождения этих этапов точность оценки эффекта ИИ повышается, однако при этом возрастают затраты ресурсов, что требует балансировки между уровнем детализации и эффективностью оценки.

Важность изучения процессов AS IS (до внедрения ИИ) и TO BE (после внедрения ИИ)

Одно из основных заблуждений заключается в том, что если модель валидирована, имеет отличные метрики качества и работает, то значит, что все отлично.

Цель внедрения моделей ИИ различается в зависимости от организации, однако с точки зрения финансовой оценки важно понимать, что успешное влияние внедренной ИИ модели на PL или любую другую денежную метрику года не всегда связано с метрикой качества этой модели. Одно из основных заблуждений, которое может быть, это ситуация, при которой единственными критериями успешности модели являются отличные метрики качества. Важно смотреть на процесс внедрения моделей ИИ широко и учитывать все потенциальные эффекты в соответствии с принципом полноты оценки.

Примеры:

Пример 1

Модель максимизирует продажи продукта X из всей полки продуктов Y.

Эффект это дельта прибыли X?

Нет, если есть другие продукты на полке, существует каннибализация с учетом ограниченного времени продающих сотрудников и замещения продуктов-аналогов.



Пример 2

Процесс сбора просроченной задолженности: у нас есть ограниченное кол-во ресурсов (СМС, звонков, сотрудников-коллекторов). Модель позволяет максимизировать объем собираемой просроченной задолженности. Значит ли это с экономической точки зрения, что модель хороша?

Поскольку нас интересует итоговый финансовый результат, нельзя смотреть только на объем просроченной задолженности, ведь итоговый результат учитывает разные компоненты прибыли продукта (комиссии, резервы, фондирование и т. д.) и соответствующие понесенные расходы на сбор. Нам важно максимизировать прибыль.



Реальность очень сложна, поэтому важно заглянуть глубже - внутрь процесса - и учесть взаимосвязи между ролями, продуктами и компонентами прибыли.

Таким образом, мы приходим к пониманию, что:

**Объект
финансовой оценки -**

изменения процессов
с применением AI модели

**Задача
финансовой оценки -**

оценить прямые и косвенные
последствия изменения или создания
процессов

Тогда нам нужно ответить на следующие вопросы:

- Как процесс устроен AS IS (бизнес-процесс без ИИ воздействия)?
- Какой драйвер в результате будет изменяться?
- Как мы будем собирать статистику по процессу до и после внедрения модели, как будет выглядеть процесс TO BE (бизнес-процесс с ИИ воздействием), какие могут быть риски?
- Какие будут эффекты финансовые и нефинансовые, если будет внедрена модель?
- Влияет ли изменение только на 1 объект управления (продукт, услугу, процесс) или же и на иные объекты управления, на которые не направлено воздействие напрямую?

Типы финансовых эффектов

Можно выделить 8 основных типов эффектов, а именно:

Тип эффекта от ИИ
инициативы

На какую финансовую метрику
оказывается влияние?

Реальное влияние на PnL

Рост доходов

Операционный доход

Снижение операционных
расходов

Операционные расходы

Высвобождение штатной
численности, ПШЕ

Операционный доход

Снижение операционных
рисков

Операционный доход /
Операционные расходы

Высвобождение
ликвидности

Операционный доход

Восстановление/
недосоздание резервов
(фактическое влияние на
кредитный риск заемщика)

Расходы на резервы

Косвенное влияние на PnL

Высвобождение резервов
(без влияние на фактический
кредитный риск заемщика) –
учетный эффект

Расходы на резервы в отчетный
период, с последующим
восстановлением резервов на
сроке жизни продукта

Снижение расходов на
капитал (за счет снижения
чистых затрат на капитал,
затрат на капитал по
резервам, в т.ч. от экономии
на объеме активов,
взвешенных с учетом риска
(RWA))

Стоимость капитала

Что такое эффект роста доходов?

Наиболее простой для понимания тип эффекта – это рост доходов. Простота данного эффекта, тем не менее, требует внимательности при оценке прямых и косвенных последствий реализуемых изменений.



*Влияет на операционный
доход Банка*

Бизнес драйверы роста дохода, можно разделить на две группы:

1. Изменение в количестве продаж
2. Изменение в маржинальности или прибыльности продаж

Причем важно отметить, что продажи и маржинальность могут рассматриваться как в разрезе отдельного продукта, так и в совокупности по нескольким продуктам в рамках сплита продаж.

За счет чего может изменяться цена и количество продаж:

- Повышение производительности / автоматизация процесса
- Запуск новых продуктов
- Изменение модели отклика в кампаниях продаж / рост эффективности коммуникаций с клиентами
- Изменение среднего чека / средней ставки / средней маржи от 1 новой продажи, а также изменение лимитов по продуктам для клиентов
- Снижение текучести сотрудников продаж и повышение эффективности подбора персонала
- Рост транзакционной активности текущих клиентов по имеющимся у клиентов продуктам
- Изменение структуры расходов на маркетинг для продвижения продуктов
- Удержание клиентов
- Расширение клиентской базы / проникновения в клиентскую базу - поиск неклиентов банка со спросом на продукты банка
- Снижение уровня отказов клиентов - например, за счет ускорения принятия решения по клиенту от внедрения модели или за счет более релевантного расчета условий по кредитам
- Повышение операционной эффективности за счёт выявления и устранения отклонений и ошибок в бизнес-процессах

☑ Что важно анализировать при эффекте роста дохода?

Важно анализировать не только прямое воздействие на целевой драйвер, но и побочные эффекты. В первую очередь это подразумевает:

Оценку эффекта каннибализации

Снижение продаж конкурирующих продуктов и услуг при росте продаж целевого продукта или услуги, или снижение продаж целевого продукта в конкурирующем канале сбыта.

Оценку эффекта кросс-продаж

Обратный эффект по отношению к эффекту каннибализации, выражающийся в росте продаж нецелевых продуктов и услуг.

Давайте рассмотрим на примере:

Пусть у нас есть модель, которая лучше оценивает склонность клиента к покупке потребительских кредитов («модель отклика»).

Пусть помимо потребительских кредитов банк продает следующие продукты: кредитные карты, ипотеки, страховки, брокерские счета, депозиты, дебетовые карты, валюту и драгоценные металлы, продукты экосистемы.

Наиболее очевидная каннибализация в такой конфигурации может возникать в продажах кредитных карт. Чем больше продается потребительских кредитов (ключевой драйвер эффекта), тем меньше у клиентов может возникать потребность в кредитных картах - и наоборот. Поскольку «конкурентные» связи между продуктами не всегда являются очевидными, наша рекомендация – проверять продажи во всем сплите. Проверка поможет точнее оценить эффект и выявить неочевидные связи в продажах.

Кросс-продажи могут возникать в продажах страховок (как сопутствующий продукт к кредиту), а также в продажах продуктов экосистемы (клиенты тратят заемные средства на покупку).

Важно!

Если связь между каннибализированным продуктом/продуктом кросс-продаж не получается логически объяснить, стоит сделать дополнительную проверку, правильно ли была сделана оценка изменения драйверов.



Причины возникновения каннибализации и кросс-продаж в данном случае:

1. Ограничения политик коммуникаций

(Например, предложение клиенту не более 3 продуктов из всей продуктовой линейки не чаще 1 раза в месяц / квартал) – все модели отклика всех продуктов по сути будут конкурировать между собой за право попасть в топ-3 при ранжировании продуктов по вероятности заинтересованности в продукте или услуге, с учетом доходности каждого продукта

2. Ограниченность времени клиентов

Клиенты имеют ограниченное время для коммуникации с Банком

3. Ограниченность времени сотрудников

Фонд времени сотрудников продаж является константой, т.е. сотрудник за один и тот же период времени может лишь по-разному распределить свое время на продажу тех или иных продуктов

Более того, сами продукты и услуги отличаются разной нормативной длительностью для осуществления продажи или услуги, таким образом увеличивая долю «длинных» операций, оставшееся время для продажи иных продуктов и услуг будет уменьшаться

Наличие кросс-канальной конкуренции

Например, увеличив продажи продуктов через Сбербанк Онлайн, может произойти кратное снижение продаж в офисах Банка (ВСП, внутренних структурных подразделениях), т.е. вместо роста общих продаж произойдет миграция продаж из одного канала продаж в другой

☑ Что такое срок жизни продукта? Зачем его учитывать в эффекте роста доходов?

В расчете эффекта роста дохода нужно всегда учитывать срок жизни продаваемых продуктов и среднюю прибыль одного проданного продукта. Разберем расчет на простом примере:

Разберем расчет на простом примере:

Предположим, что у нас есть 5 продуктов. Был проведен А/Б тест, по 2 из 5 продуктов было выявлено значимое изменение в конверсии (по первому +10%, а по второму -2%). Если по первому продукту срок жизни 1 месяц, а по второму – 10 лет, то общий эффект от инициативы будет отрицательным (при сопоставимых значениях средней прибыли продуктов), хотя, казалось бы, прирост конверсии по первому продукту больше, чем падение конверсии по второму. Дополнительное усложнение расчета возникает, если значения средней прибыли продуктов 1 и 2 сильно различаются.

Поэтому при подведении итогов важно учитывать прибыль, срок жизни и конверсии по всем продуктам на полке продаж.

Расчет эффекта в таком случае будет выглядеть повинтажно, т.е. с построением потоков прибыли с учетом момента возникновения продаж, закрытием старых договоров и до конца срока жизни продукта. Такой формат позволяет наиболее точно оценить эффект на PnL банка, учитывать изменения в отклике, структуре и рентабельности продаж от месяца к месяцу.

Как альтернативу продуктовому подходу со сроком жизни – возможно использовать метрики клиентского LTV (CLTV), если такая метрика доступна и корректно рассчитывает ценность клиента в компании. LTV возможно использовать для оценки внутри компании, но для внешнего репортинга мы рекомендуем использовать метрики в терминах PnL.

Что такое эффект снижения операционных расходов?

Эффект от снижения операционных расходов аналогичен росту доходов, только с обратным знаком: чем больше мы сокращаем расходы, тем больший эффект наблюдаем. Драйверами эффекта в данном случае являются количество совершаемых расходных операций и их стоимость – на каждый компонент мы можем повлиять, принимая решение или выявляя потенциальные улучшения в текущих процессах с применением ИИ инструмента.

Пример:

Внедряется ИИ маршрутизация по картам геолокации курьеров/инкассации/специалистов прямых продаж, таким образом оптимизируются драйвер – затраты на топливо для выполнения плана доставок/инкассации/продаж. Уменьшение затрат является финансовым эффектом от инициативы.

Если при мониторинге мы обнаружим значительное изменение целевого значения, это будет сигналом к проверке: сохраняется ли эффект, насколько он выражен и можем ли мы на этом основании делать выводы. Это можно определить, например, проанализировав естественную историческую волатильность показателя и проведя тест на наличие стат.значимого отклонения.

Что такое эффект высвобождения штатной численности (эффект ПШЕ)? Какие примеры эффектов ПШЕ есть?

Эффект от высвобождения штатной численности измеряется в терминах количества сотрудников / ПШЕ (полных штатных единиц). Соответственно, ключевыми драйверами в данном случае являются количество совершаемых операций и их нормативная длительность.

Высвобождение численности подразделяется на два типа:

1. Фактические (реальные) сокращения сотрудников
2. Исключение потребности в дополнительных сотрудниках (ненайм)

Важно, чтобы эффективность ИИ в сравнении с «ручным» трудом была доказана при помощи методов пилотирования или хронометража / замеров – для проверки эффективности ИИ решений на практике. Расчет должен быть подтвержден владельцем моделей численности организации и найти отражение в бизнес-плане компании, иначе эффект будет виртуальным и компания не сможет увеличить свою прибыль в реальности.

1. Эффект реального сокращения численности

Данный тип эффекта подразумевает воспроизведение части функционала сотрудников за счёт ИИ алгоритмов, в результате чего происходят автоматизация функционала и оптимизация нормативной численности, т.е. сокращаем реальное количество сотрудников.

При расчете финансового эффекта от фактического сокращения, помимо очевидного прямого эффекта от сокращения расходов на заработную плату, нужно учесть все сопутствующие расходы и доходы, которые свойственны рассматриваемому процессу. Например, необходимо учесть выплаты сотруднику дополнительных денежных средств в случае досрочного прекращения трудового договора со стороны работодателя.

При переходе от ручного труда к ИИ решениям могут возникать эффекты на иные показатели эффективности оптимизируемого процесса, и эта связь с ключевым драйвером, улучшаемой метрикой, не всегда очевидна, но требует внимания, следуя принципу полноты.

Например, инициатива автоматизации принятия решений может уменьшить количество операционных дефектов, стоимость которых отразится на PnL. Возможен сценарий, когда автоматизация может привести и к увеличению объемов продаж или уменьшению операционных издержек или же ухудшить эффективность продаж при полной автоматизации функционала. Многие эффекты, связанные с замещением ручного труда, являются «многослойным», с большим количеством драйверов внутри бизнес процесса, которые необходимо измерять до и после изменений и мониторить в динамике.

2. Эффект исключения потребности в дополнительных сотрудниках (ненайм)

Эффект возникает, когда при росте бизнес-драйвера численности (росте количества клиентов, которых нужно обслуживать, росте заявок, которые нужно обрабатывать, и т.д.) увеличения численности не происходит за счет внедрения ИИ модели. Такого рода модель может, например, увеличивать эффективность работы текущего штата сотрудников, что позволяет им обрабатывать и принимать решения на большем объеме входящего потока задач с меньшими издержками по времени.

Важным моментом в данном случае является фокус на том, как считается эффект ненайма – это должно быть объективной оценкой на основании драйверов процесса, количества выполняемых операций и их длительности. При этом функционал должен быть нормативно обязательным для подразделения.

Рассмотрим конкретный пример, покрывающий одновременно эффект реального сокращения ПШЕ и эффект ненайма:

В качестве бейзлайна инициативы используются ИИ модели, внедренные в процесс принятия решения по кредитным заявкам. Для проверки того, что автоматизация проводилась именно за счет ИИ, была сделана кросс-проверка функций, выполняемых штатной численностью, и драйверов процесса, на которые влияли указанные ИИ модели.

За счет внедрения ИИ, коэффициент автоматизации обработки заявок был повышен с 60% до 90%. В начале входящий поток заявок был 100 шт. 40% заявок обрабатывалось в ручном режиме, на что требовалось 4 сотрудника (производительность – 10 заявок на человека).

Год	0	1
Коэффициент автоматизации	60%	90%
Количество входящих заявок для обработки	100	200
Численность ПШЕ	4	2
-	-	-
Эффекты	-	-
Реальные сокращения	-	2
Ненайм	-	4

При одновременно повышении входящего потока заявок со 100 штук до 200 и коэффициента автоматизации с 60% до 90% в следующем году потребовалось 2 сотрудника на ручную обработку. Размер эффекта реального сокращения и ненайма в терминах ПШЕ:

- Реальное сокращение посчитать довольно просто:

Поскольку мы знаем (проверили это заранее), что на процесс не влияло ничто, кроме рассматриваемой ИИ инициативы, фактические сокращения равны численность на начало периода минус численность на конец периода. Наблюдается эффект в 2 ПШЕ.

- Ненайм в такой конфигурации рассчитывается несколько сложнее.

Представим, что ИИ модель не внедряли бы, при этом на процесс не оказывалось никакого стороннего влияния (единственное, что могло меняться от года к году – это улучшение ИИ модели, дообучение/включение новых параметров). В таком случае, коэффициент автоматизации в последующих годах остался бы на уровне 60%, или потребовалось бы 8 человек для ручных проверок, 4 из которых уже работали изначально. Таким образом, пришлось бы нанять дополнительно 4 сотрудника в штат. Эффект ненайма получается 4 ПШЕ.

Что такое эффект на снижение операционных рисков, приводящих к недополучению прибыли / росту операционных расходов?

Механизм влияния операционных рисков на PnL аналогичен эффектам дохода и эффектам операционных расходов, в зависимости от полученного знака.

Драйверы, через которые данный эффект оказывает влияние на PnL – это:

- **Операционные дефекты**
- **Фрод / мошенничество**

Как по дефектам, так и по фроду, либо должна вестись статистика, чтобы можно было оценить средний «вклад» показателя в PnL, либо необходимо провести А/Б тест, проконтролировав все изменяемые переменные, вычленив необходимый эффект в целевой группе относительно контрольной группы.

По выбранным драйверам важно всегда устанавливать мониторинг. Необходимо оценивать, насколько изменяется объем фрода/количество дефектов, выявляемых ИИ моделью, в общем количестве фрода/дефектов.

Если ИИ модель год к году стала выявлять больше фрода/дефектов, при этом сама модель не дорабатывалась, а общее количество фрода/дефектов стабильно или вовсе снижается – это является маркером для дополнительной проверки, нет ли деградации используемого алгоритма. Возможно, оценка искажена или ошибка кроется в расчетах изменения драйвера. В таком случае требуется переоценка, а возможно, и проведение дополнительного пилота для корректировки эффекта. Также возможным исходом в рамках мониторинга стабильности работы модели может быть решение о необходимости переработки модели.

Разберем простой пример из реальной бизнес-практики по эффекту операционных дефектов.

NB! Обратите также внимание, что часто эффекты комплексны - это несколько эффектов в одной инициативе, и данный кейс не является исключением.

В бизнес-процесс внедряется ИИ модель, которая прогнозирует вероятность совершения сотрудником ВСП дефекта по операции. Сам бизнес-процесс – это проведение проверок сотрудников банковских офисов их руководителем / аудитором на предмет наличия операционных дефектов в совершаемых операциях. Предположим, ИИ модель показала свою эффективность.

Какой от этого финансовый эффект может возникнуть?

- Высвобождение численности аудиторов / руководителей, которые должны совершать меньшее количество проверок.
- Сокращение издержек при более точечных проверках по модели количества операционных дефектов, стоимость которых известна по собранной статистике.

Что такое эффект на высвобождение ликвидности?

Данный тип эффекта возникает, когда ИИ модель высвобождает объем денежных средств, ликвидности, которую нужно держать, например, в рамках норматива ликвидности или СДО денежных средств в кассовых центрах или банкоматах. Высвобождаемую ликвидность можно размещать по ставке Overnight, что обеспечит дополнительную прибыль компании.

Драйвером эффекта в данном случае будет выступать объем ликвидности, на который влияет ИИ. Ставка, под которую размещаются денежные средства, является также изменяемым фактором, но она экзогенна по отношению к ИИ инициативе и нужна для мониторинга эффекта, поскольку от нее зависит доходность. Важна она еще и потому что, если единственной альтернативой размещения высвобожденной ликвидности будет размещение ее под отрицательную ставку, не важно, какой объем нам позволяет сэкономить ИИ, финансовый эффект такого действия будет отрицательным.

Также важно анализировать ошибки первого и второго рода в решениях, связанных с управлением ликвидностью: если мы ошибемся с оценкой свободной ликвидности для размещения ее на рынке, может реализоваться риск неспособности выполнить договорные отношения с клиентами из-за дефицита ликвидности, в результате чего компании придется привлекать средства по более высокой ставке на рынке.

Что такое эффекты на резервы? Что важно знать при оценке этих эффектов?

При расчете эффекта резервов необходимо корректно определить тип резервов, драйверы изменения резервов, понять, постоянный или временный эффект мы имеем, а также учесть принцип неттирования.

Важно помнить, что существует два типа резервов:

Первый тип – это регуляторные резервы.

Это пруденциальные резервы, устанавливаемые Банком России, которые коммерческий Банк обязан перечислять на корреспондентский счет в Банке России.

Согласно российскому законодательству, Банки должны формировать Обязательные Резервы, Резервы под возможные потери по ссудам, Резервы под возможные потери и Резервы под обесценение ценных бумаг. Первые три типа наиболее часто встречаются при оценке ИИ эффектов. Кратко про каждый из них.

- **Обязательные резервы** - это часть суммы депозитов Банка, которую необходимо хранить на счете в Банке России.
- **Резервы на возможные потери по ссудам** (их сокращенно называют РВПС) формируются при потере ссудной стоимости, из-за ненадлежащего исполнения обязательств заемщиком. Например, клиент, которому был выдан кредит, задерживает очередной платеж на 40 дней. Банку необходимо сформировать

дополнительный резерв по такой ссуде.

- Резервы по возможным потерям (сокращенно – РВП) формируются по балансовым активам с риском потерь (за исключением тех активов, по которым сформированы РВПС), по условным обязательствам и прочим потерям

Второй тип резервов – резервы по стандартам отчетности.

Простыми словами, это фонд, создаваемый организацией для покрытия своих убытков и отражаемый в отчетности. Два основных стандарта отчетности: международные стандарты финансовой отчетности (МСФО) и российские стандарты (РСБУ). Резервы по МСФО и по РСБУ не равны, так как в методологиях содержатся различные правила учета отдельных статей баланса.

Можно выделить три типа эффектов на резервы:

*Временное сокращение /
увеличение резервов*

*Увеличение / уменьшение
срока сбора просроченной
задолженности*

*Изменение объема сбора
просроченной
задолженности*

Давайте разберем каждый из типов подробнее на примерах.

1. Временное сокращение / увеличение резервов

Пусть разработанный ИИ-алгоритм проверки базы данных просроченных кредитов показал, что начисление резерва в размере 25% по ряду займов на сумму 400 млн руб. не требуется, поскольку они обеспечены залогом. По таким займам норма резервирования, например, должна быть 0%. То есть мы начисляем резерв 25%, а можно начислять 0%. И мы исправляем эту ошибку.

Исправление ошибок позволило сократить резервы на 100 млн руб. ($400 \times 25\%$). Пусть ошибочно созданные резервы были бы восстановлены без модели через 180 дней. 180 дней – это длительность эффекта, мы имеем временный эффект. Эффект, учитывая стоимость денег во времени, составит 2,5 млн руб.

При подтверждении таких эффектов надо помнить, что эффект не признается по кейсам, когда изменяется политика Банка по управлению резервами или нормы регулятора по изменению учета РВПС. Так как изменение резервов в таких случаях не является последствием внедрения ИИ-моделей. Модель в данном случае нужна, чтобы выполнить изменившиеся обязательные требования.

2. Увеличение/уменьшение срока сбора просроченной задолженности

Ускорение процесса сбора проблемной задолженности приводит к более раннему поступлению денежных средств в Банк. Пусть ИИ-алгоритм позволил сократить срок взыскания проблемной задолженности на 1 месяц. Объем взыскиваемой проблемной задолженности – 8 млн руб., покрытие резервами – 100%. Если средняя стоимость капитала составляет 20%, то эффект за год составит 133 тыс. руб. ($8 \times 0,2 = 1,6$ млн. $1,6$ млн./12 мес. = 133 тыс. за 1 месяц)

3. Изменение объема сбора просроченной задолженности

При увеличении сбора безнадежной задолженности эффект представляет собой всю сумму дополнительно собранной задолженности (не забываем, что за вычетом лучшего альтернативного варианта работы с безнадежной задолженностью).

Например, коллекторы в результате внедрения модели научились собирать проблемную задолженность, которая ранее признавалась безнадежной и продавалась по цессии за 10% номинала. Объем такой дополнительно собранной задолженности – 1 млрд руб., покрытие резервами – 100%. Поскольку взыскание такой безнадежной задолженности прямо попадает в капитал Банка, то эффект составит 900 млн руб.

☒ Принцип неттирования в эффектах на резервы и почему его важно учитывать в эффектах на резервы

Эффекты изменения резервов специфичны своим изменением во времени. В один период времени резерв формируется, что отражается в отчетности увеличением резерва, в другой период времени происходит высвобождение резервов, что также находит отражение в отчетности. Поэтому нужно смотреть life-time эффект, не ограничиваясь одним отчетным периодом. Для некоторых случаев влияния на резервы, без прямого влияния на кредитный риск, life-time эффект может вырождаться в 0. Такой случай был назван принципом неттирования.

Принцип неттирования для понимания лучше разобрать на примере.

Пусть есть ИИ-инициатива, внедрение которой приводит к тому, что меняется временная структура создания и высвобождения резервов. Например, модель рекомендует привлекать на продукт «Кредит» клиентов, отличных от тех, которые привлекались бы в процессе без ИИ. Для простоты будем рассматривать только клиентов, которые приобрели кредит в 2020 году.

Клиенты, привлекаемые с помощью ИИ (это процесс TO BE), оцениваются как более благонадежные, и по ним формируются резервы в размере 20 млн. руб., в то время как по клиентам, привлекаемым без ИИ (это процесс AS IS), сформированы резервы размером в 32 млн. руб. Положительный эффект

составил 12 млн. руб.

В 2021 году резервы снизились в AS IS процессе на 12,8 млн. руб. до 19,2 млн. руб., в TO BE – на 8 млн. руб. до 12 млн. руб. Возник отрицательный эффект в PnL 2021 года равный -4,8 млн. руб.

В 2022 году эффект в данном кейсе также отрицательный.

Таким образом, эффект резервов от года к году может меняться не только в объеме, но и по знаку (с + на – и наоборот), поэтому важно рассматривать такие эффекты life-time.

	Дата отчета		
	31.12.20	31.12.2021	31.12.2022
Портфель 2020 на дату	100	60	0
Резерв до	32	19,2	0
Резерв после	20	12	0
Дельта резервы	12	7,2	0
Фин. рез. Год-Год AS IS		12,8	19,2
Фин. рез. Год-Год TO BE	12	8	12
Дельта фин. рез.	12	4,8	-7,2

Что такое эффект снижения расходов на капитал?

Говоря про эффект изменения расходов на капитал, стоит начать с того, что расходы на капитал состоят из трех элементов: чистые затраты на капитал, называемые ЧЗК, затраты капитала на резервы и затраты на фондирование.

Поэтому эффект в этой части может возникать через изменение Чистых Затрат на Капитал (ЧЗК), через изменение затрат капитала на резервы и через изменение активов, взвешенных с учетом риска (RWA).

Рассмотрим пример:

Пусть есть ИИ-инициатива, внедрение которой приводит к тому, что меняется временная структура создания и высвобождения резервов. Например, модель рекомендует привлекать на продукт «Кредит» клиентов, отличных от тех, которые привлекались бы в процессе без ИИ.

Предположим, что в процессе TO BE клиенты имели значительные задержки в совершении платежей, и в Банке наблюдался высокий уровень просроченной задолженности.

Это значит, что в процессах AS IS и TO BE отличаются значения показателя RWA – активов, взвешенных с учетом риска.

Теперь разберемся с показателем ЧЗК. Чистые затраты на капитал - это объем задолженности с учетом процентов, штрафов, пени, неустоек за вычетом регуляторных резервов и скорректированный на стоимость рублевого акционерного капитала, норматив достаточности капитала, риск-вес активов и количество дней в году.

$$\text{ЧЗК} = ((\text{ОСЗ} + \text{Проценты}_i + \text{ШПН}) - \text{Регуляторные резервы}) \times (\text{CoE}/365) \times \text{H1} \times \text{RWA}_i$$

где ОСЗ - остаток ссудной задолженности, ШПН - штрафы, пени и неустойки, CoE - стоимость капитала, H1 - норматив достаточности собственного капитала (H1.0).

Изменение компонентов объема задолженности и RWA приводят к изменению Чистых затрат на капитал.

Кроме того, значение риск-веса участвует также в расчете капитала:

$$\text{Собственный капитал}_i = ((\text{ОСЗ} + \text{Проценты}_i + \text{ШПН}_i) - \text{Регуляторные резервы}_i) \times H1 \times RWA_i$$

А затраты на фондирование, в свою очередь, зависят от объема капитала Банка:

$$\text{Фондирование}_i = ((\text{ОСЗ} + \text{Проценты}_i + \text{ШПН}_i) - \text{Регуляторные резервы}_i - \text{Собственный капитал}_i) \times (ETC/365),$$

где ETC - единая трансфертная ставка.

Финансовый эффект от реализации модели = изменение веса и значений компонент стоимости капитала по процессу урегулирования проблемной задолженности

В завершении темы про Резервы и Капитал, нужно отметить, что мониторинг подобных эффектов крайне важен. Необходимо контролировать изменение драйверов эффекта (например, стабильности уровня сбора задолженности, расходов на сбор задолженности, доля кредитов, уходящих в дефолт) и в соответствии с изменениями корректировать расчет эффекта. Также стоит обращать внимание на эффекты второго порядка, так как существует тесная связь между показателями, участвующими в расчете финансового результата.

Что такое инкремент?

Инкремент – изменение драйвера, метрики, финансового эффекта, возникшее в результате конкретного воздействия на процесс по итогу внедрения конкретной модели ИИ / инициативы с ИИ.

Понятие инкремента неразрывно связано с понятием бейзлайна (процесс или состояние процесса, относительно которого измеряется инкремент), а также с принципом 4. «Принцип связи с инициативой».

Важно понимать, что бейзлайн для финансовой оценки / для оценки экономического эффекта от ИИ инициативы не всегда равен процессу AS IS. Бейзлайн для оценки представляет собой **лучшую ближайшую альтернативу процессу TO BE**.

- Если процесс AS IS представляет собой лучшую ближайшую альтернативу, то для получения оценки достаточно сравнить процессы AS IS и TO BE между собой
- Если создается новый продукт / процесс – анализируются альтернативные издержки ближайшей альтернативы (например, BI⁶ решение) для достижения аналогичной цели, и оценивается ИИ инкремент относительно выработанных BI инструментов. Тут необходимо также руководствоваться принципом материальности и не создавать слишком дорогую, но точную BI альтернативу.

Крайне важно понимать, что эффект от внедрения модели измеряется именно против лучшей альтернативы, в противном случае в финансовую оценку дохода от ИИ будет замешана оценка дохода от изменения процесса с автоматизации, но без ИИ.

6 BI - Business Intelligence или бизнес-аналитика

Примеры:

Исходящий обзвон клиентов ИИ-ботом для напоминания о брошенной заявке

- Изменения: добавление нового процесса и внедрение ИИ в процесс
- Бейзлайн некорректно сравнивать относительно отсутствия звонка, правильным бейзлайном может быть лучший вариант из:
 - Звонок оператором
 - Звонок с предзаписанной аудиодорожкой некорректно сравнивать относительно отсутствия звонка

Против ИИ модели рекомендательной системы какого-либо продукта в качестве наилучшей альтернативы могут выступать сформулированные BI правила рекомендации такого продукта.

Мы должны понимать, что если бы не было ИИ, то цель могла бы быть достигнута другими аналитическими средствами (BI правилами, экспертами, которые консультировали клиентских менеджеров и т.д.).

Внедряем новый продукт – сервис новостного мониторинга.

- Модель в моменте анализирует контекст статей и заголовков на ресурсах, агрегирующих новости, на предмет выявления сигналов по потенциальному изменению риска клиентов-заемщиков сегменте КИБ⁷
- Бейзлайн: так как модель в данном случае неотделима от процесса (невозможно было бы осмысленно анализировать такое количество новостных статей в единицу времени), то оценку инкремента необходимо производить с работой процесса до внедрения сервиса новостного мониторинга.

Может также возникнуть проблемы с выбором бейзлайна при сложной, составной инициативе и долгосрочных проектах.

Пример:

Возьмем кредитно-скоринговую модель, экономический эффект за год N будем считать как:

Прирост (в руб.) на кредитную заявку относительно бейзлайна * Кол-во кредитных заявок за год N

Кейс 1

Предположим первую версию собственной модели разработали в начале 2023 года, до этого использовали внешние скоринговые баллы.

В таком случае, эффект от ИИ за 2023 год будет считаться по формуле:

Прирост (в руб.) на кредитную заявку относительно внешних скоринговых баллов * количество кредитных заявок за 2023

Кейс 2

В начале 2024 года разработали новую версию кредитно-скоринговой модели, которая лучше, чем версия 2023 года. Как считать экономический эффект от модели за 2024 год:

Вариант 1:

Прирост (в руб.) на кредитную заявку относительно внешних скоринговых баллов * количество кредитных заявок за 2024

Вариант 2:

Прирост (в руб.) на кредитную заявку относительно модели v2023 * количество кредитных заявок за 2024

Сравнение двух способов:

В первом варианте мы считаем общий экономический эффект от ИИ, который банк фактически получил за 2024 год. Однако, часть данного эффекта была достигнута за счет разработок сделанных в 2023 году, соответственно расходы на эти разработки были учтены в 2023 году.

В рамках методологии мы хотим считать полный экономический эффект, однако если в управленческих вопросах опираться только на данную метрику, могут быть приняты ошибочные решения.

Пример:

В 2020 году мы наняли команду, которая стоит 100 млн руб в год. Команда сделала модель, которая генерирует 150 млн руб. в год относительно VI-бенчмарка. Предположим, в следующие 3 года, команда выкатывала минорные изменения, которые приносили по 10 млн прироста в год относительно предыдущей версии модели, тогда экономический эффект от их модели будет:

- 2021: 160 млн
- 2022: 170 млн
- 2023: 180 млн

Если сравнивать затраты на команду с экономическим эффектом от их модели, то выглядит, что содержание команды довольно выгодно для компании, однако инкремент их работы за 2021-2023 гораздо меньше, чем затраты на команду за тот же период.

Эффект от сложносоставных инициатив и долгосрочных проектов

Базовой рекомендацией методики является расчет полного экономического эффекта

относительно лучшей альтернативы без ML. Однако в долгосрочных проектах бывают проблемы с поддержанием изначального бейзлайна. В этих случаях часто прибегают к оценке относительно предыдущей версии модели, а итоговый экономический эффект за период считают как сумму инкрементальных изменений.

Важно учитывать, что сумма инкрементальных изменений не всегда равна совокупному эффекту относительно бейзлайна, так как может присутствовать каннибализация (или, реже, синергия) между несколькими версиями обновлений.

Интуитивно это выражается так:

Инкремент (V1 vs V3) $\neq$ Инкремент (V1 vs V2) + Инкремент (V2 vs V3)

В случае подхода с суммированием инкрементов рекомендуется калибровать оценку через замер инкремента относительно версии модели на несколько поколений назад.

Практика: раз в год проводить «обратный» эксперимент для замера эффекта последней версии модели относительно версии, существовавшей на начало года (или на k поколений назад), и использовать эту калибровку для корректировки суммы инкрементов.

☑ Мониторинг эффектов

Чтобы следовать принципу измеримости, а также иметь возможность в любой момент жизни ИИ инициативы ответить на вопрос: какую экономическую и финансовую выгоду она приносит, – в рамках процесса оценки необходимо всегда выставлять регулярный мониторинг оцененного эффекта.

Регулярный мониторинг выставляется как на метрики качества модели (Precision, Recall, GINI, MAPE, MAE, т.д.), так и на финансовые и операционные драйверы оценки.

☑ Метрики качества

Все модели деградируют со временем, если их не дообучать, из-за изменения входного потока данных и изменения закономерностей.

Для разного типа моделей мониторятся разные метрики точности. Какие именно - зависит от типа ML задачи и Критичности бизнес-процесса.

Как правило, чем выше цена ошибки при работе модели (например, выдача кредита, блокировка транзакции) тем ниже толерантность к допустимому снижению метрик модели.

Модель обновляется, когда мы видим снижение метрик ниже допустимого порога. Перспективным выглядит автомониторинг моделей; в отсутствие автомониторинга – мониторинг качества производится с определенной регулярностью в зависимости от критичности процесса, например модели высокой степени значимости мониторятся чаще, чем модели низкой степени значимости.

Рекомендуемая частота замеров качества модели – раз в месяц.

Минимально допустимая в зависимости от степени значимости моделей:

Критически значимые (напр., значительный фин. эффект или затрагивает критически значимые процессы)	Высокой степени значимости (напр., средний фин. эффект или затрагивает значимые процессы)	Средней степени значимости (напр., низкий фин. эффект или затрагивает процессы средней значимости)	Низкой степени значимости (напр., без фин. эффекта и не имеющие существенного влияния на важные процессы)
Раз в 3 месяца	Раз в полгода	Раз в год	Не проводится

Финансовые и операционные метрики

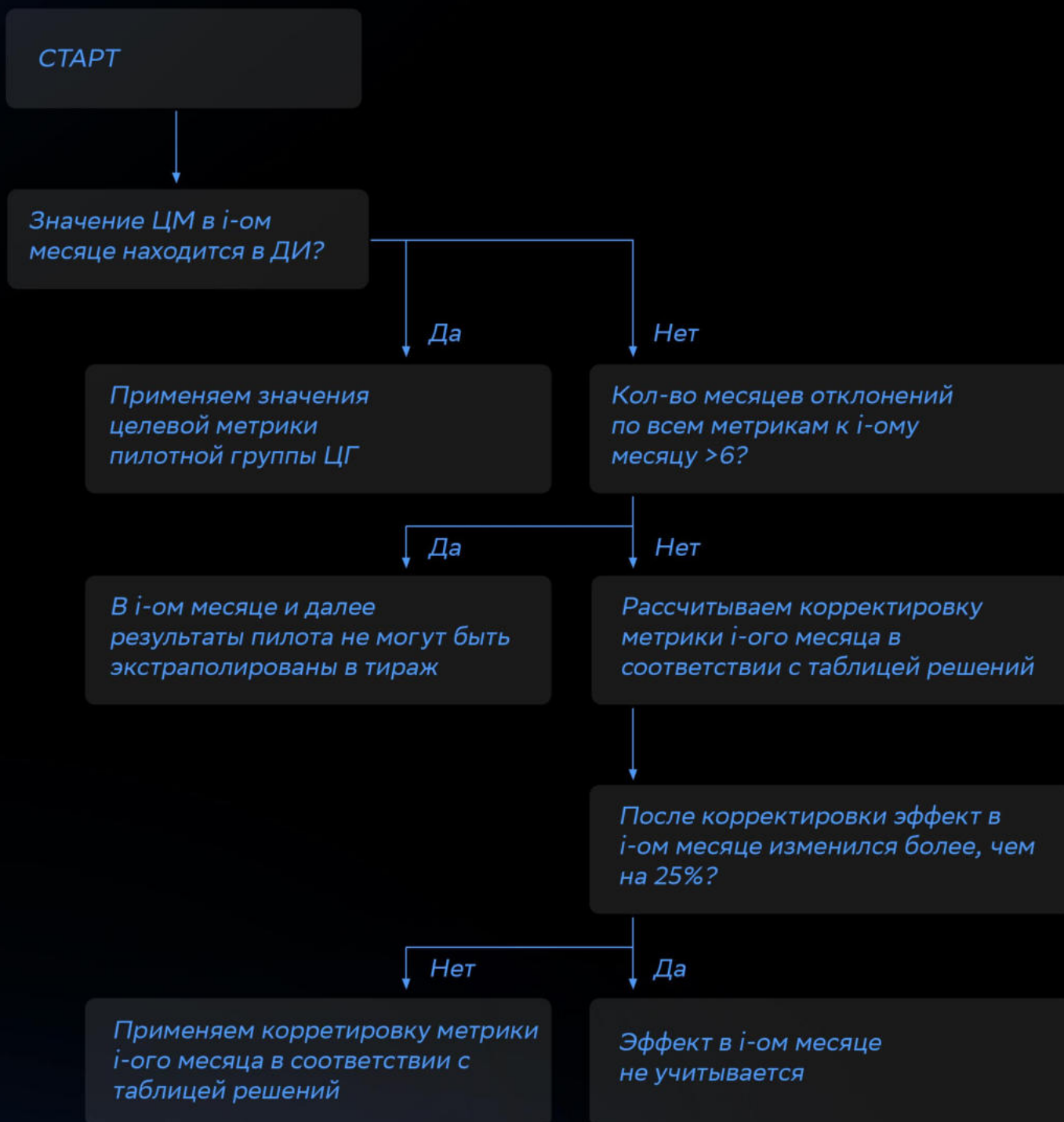
Что касается финансовых и операционных метрик, частота и метод мониторинга сильно зависят от типа эффекта рассматриваемой ИИ инициативы, а также наличия или отсутствия постоянной контрольной группы.

Для случая, когда постоянная КГ (СКГ) присутствует, необходимо проводить замер инкремента драйверов относительно этой контрольной групп; при этом необходимо проверить СКГ на предмет вымывания/эрозии.

- Не реже раз в 6 месяцев с момента первой оценки для эффектов на рост дохода, снижение операционных расходов, снижение операционных рисков и высвобождение ликвидности рекомендуется проводить мониторинг.
- Не реже раза в 12 месяцев для эффектов на высвобождение штатной численности, восстановление/ недосоздание резервов, а также снижения расходов на капитал рекомендуется проводить мониторинг в конце каждого календарного/финансового года.

Для случая, когда постоянная КГ (СКГ) отсутствует, рекомендуется проводить ежемесячную проверку эффекта и сравнивать его с доверительным интервалом, полученным на этапе проведения А/Б теста.

На примере:



Чистый (net) vs Валовый (gross)

В оценку ИИ-проектов необходимо включать все затраты на создание и поддержку. Считаем, что чистый (net) подход самый правильный для расчета фин. эффекта от ИИ.

Преимущества чистого подхода:

- Полный учет инвестиций и реалистичная оценка прибыльности (ROI)
- Объективное сравнение проектов
- Эффективное управление рисками

Учитываемые затраты.

При расчете чистого эффекта учитываются все расходы, относимые к созданию и поддержке ИИ:

- Фонд оплаты труда (ФОТ)
- Серверная инфраструктура
- Внешние расходы:

Все внешние закупки - лицензии / консалтинг / услуги / поддержки / разработка с использованием персонала вне штата/ внешняя (аутсорс) разработка.

1. ФОТ

Все расходы связанные с человеческим капиталом:

- валовый оклад, налоги (помимо тех, что учтены в валовом окладе), соц взносы, все компенсации, районные надбавки, отпускные, бонусы, премии
- вмененные расходы (аллокация стоимости офисного места, техники используемой сотрудником); необходимость учета вмененных расходов определяется финансовой службой на основании предоставленных обоснований.

Оценивается по PnL (амортизация нематериального актива), не с точки зрения движения денежных средств.

Большая часть разработки ИИ решений не ставится как НМА, т.к. сложно верно определить СПИ из-за динамически меняющейся отрасли.

Аллокация сотрудников: если сотрудник был привлечен в разработку ИИ-продукта или процесса, то затраты на него учитываются в затратах на ИИ.

2. Инфраструктура

Расходы учитываются с позиции PnL-отчета, а не с точки зрения движения денежных средств

Делятся на 2 категории - инфраструктура, закупленная только под инициативы ИИ (например GPU), и общебанковская инфраструктура, которой пользуются команды для создания ИИ продуктов на основе внутренних тарифных моделей

Первая категория полностью относится к "инвестиции в ИИ".

В расход берут амортизацию серверов, часть расходов датацентров, аллоцированных на данные сервера, команду (тех .поддержка оборудования / сети), необходимые лицензии для поддержания данных серверов, если есть

Вторая категория аллоцируется на сущность "инвестиции в ИИ" пропорцией из метрик потребления общебанковских платформ (платформа виртуализации, базы данных, логирование, аналитические платформы, инференс и т.д.).

Важно отметить, что это аллокация фактических затрат, а не биллинг. Из-за этого стоимость единицы ресурса не фиксированная, а зависит не только от общей стоимости платформы, но и от общего потребления / утилизации по компании инфраструктурных платформ

3. Внешние расходы

Все внешние закупки - лицензии / консалтинг / услуги / поддержки / разработка с использованием персонала вне штата/ внешняя (аутсорс) разработка

Важно определить драйвер аллокации внешних расходов, например

- Приписывание на команду, запросившую расход
- На основе использованных токенов/карт/иных счетных и отдельных параметров
- На основе внутренних тарифных моделей

Гранулярность оценки каждого из пункта выше - 1 раз в месяц.

Какими метриками можно оценить фин. эффект от ИИ?

Набор метрик для оценки финансовой эффективности в зависимости от целей и задач

Потенциальные эффекты от ИИ:

Внедрение технологий искусственного интеллекта (ИИ) в бизнес-процессы может приносить различного рода финансовые эффекты, которые важно учитывать при оценке их общей эффективности. Эти эффекты можно условно разделить на прямые и косвенные, а также стратегические. Прямой финансовый эффект проявляется в виде конкретной экономии или дополнительного дохода — он может сразу отразиться на прибыли (PnL) текущего года, например, за счет снижения издержек на поддержку клиентов благодаря чат-ботам. В то же время некоторые прямые эффекты, такие как формирование дополнительного будущего денежного потока (например, за счет выхода на новые сегменты рынка с ИИ-продуктами), могут не повлиять на PnL текущего периода, но станут источником роста выручки и прибыли в последующие годы. Помимо прямых, также возможны косвенные финансовые эффекты и стратегические эффекты, которые, хотя и не всегда измеряются напрямую через PnL, вносят вклад в устойчивость и развитие бизнеса в среднесрочной и долгосрочной перспективе.

Таким образом, можно выделить следующие эффекты от внедрения ИИ:

- Прямые финансовые с влиянием на PnL (например, доп. выручка и снижение затрат)
- Прямые финансовые без влияния на PnL текущего года (доп. будущий денежный поток - например, NPV)
- Косвенные финансовые эффекты с влиянием на PnL (снижение расходов на капитал)
- Сэкономленные деньги (например, «не найм» сотрудников) (виртуальный/косвенный PnL)
- Косвенные стратегические эффекты (CSI/NPS/CSAT, Loss prevention и т.д.)

Предлагаемые метрики:

Потенци- альные эффекты	Метрика	Внешние задачи (PR/IR)	Внутренние задачи				
			Приоритизация		Оценка эффективности инвестиций на уровне		Учет в фин. эффекте компаний
			Инициативы	Портфеля инициатив	Инициативы	Портфеля инициатив	
Доп. выручка и снижение затрат	PnL отчет- ного года	△			△	△	△
	PnL на 12 мес. вперед		△	△	△	△	
Доп. будущий денежный поток	NPV		△	△	△	△	
	CLTV		△	△	△	△	
Спасение средств клиента	Loss prevention		△		△		
Косвенные эффекты от улучшен- ного поль- зовательс- кого / клиентс- кого опыта	CSI/NPS/ CSAT		△		△		
Рост скорости / производи- тельности	Time-to- XXX		△		△		
Сэконом- ленные деньги (например, «не найм» сотрудни- ков)	Виртуаль- ный/ косвенный PnL		△	△	△	△	

Для эффективной оценки и коммуникации эффектов от внедрения ИИ-инициатив рекомендуется использовать структурированный подход, который учитывает разные типы бизнес-эффектов и сопоставляет их с соответствующими метриками в зависимости от цели — будь то внутреннее принятие решений или внешняя отчетность.

Для каждого эффекта можно сопоставить конкретные метрики, позволяющие его измерить и обосновать.

Так, для оценки влияния на текущую прибыль следует использовать PnL отчетного года или PnL на 12 месяцев вперед — эти метрики наиболее релевантны для анализа краткосрочной эффективности. При этом первая позволяет привязываться к циклу

бизнес-планирования, а вторая - следить за тем, какие эффекты будут от портфеля на горизонте ближайшего года и оперативно реагировать на изменения. Для оценки долгосрочной ценности проектов рекомендуется применять NPV и CLTV (при этом CLTV может быть также рассчитана для текущей выручки), особенно когда эффект проявляется не сразу, а формирует будущие денежные потоки. Метрика Loss prevention полезна для количественной оценки предотвращенных потерь, например, за счет снижения мошенничества или оттока клиентов. В свою очередь, показатели CSI, NPS, CSAT помогают отслеживать косвенные эффекты, связанные с качеством взаимодействия с клиентами, а Time-to-XXX — фиксировать прирост скорости выполнения процессов. Экономия, достигнутую за счет автоматизации и отказа от найма, можно оценивать через виртуальный PnL.

Для внешнего репортинга рекомендуем использовать PnL отчетного года как метрику, наиболее прозрачно отражающую влияние ИИ на финансовый результат компании. Она соответствует стандартам финансовой отчетности, легко интерпретируется и позволяет избежать субъективных оценок. Все прочие метрики, даже если они значимы для внутреннего анализа, могут быть восприняты как недостаточно надежные или сложные для верификации — поэтому в коммуникации с инвесторами, аналитиками и другими внешними сторонами мы рекомендуем опираться исключительно на PnL отчетного года.

На этапе массовой приоритизации инициатив предлагаем использовать одну из следующих метрик в зависимости от целей приоритизации, волатильности рынка и характера проектов:

- PnL на 12 месяцев вперед — для инициатив с ближайшим финансовым эффектом удобно использовать для приоритизации при массовом запуске большого количества проектов и волатильном рынке (где сложно прогнозировать NPV);
- NPV — для капиталоемких или долгосрочных проектов;
- CLTV — для инициатив, направленных на удержание или монетизацию клиентов, в случае, если возможно посчитать влияние инициативы на CLTV;
- Виртуальный/косвенный PnL — для проектов, где эффект выражается в экономии на ресурсах, например, за счет автоматизации функций.

В случае, если поступает большое количество инициатив, чей результат сложно перевести в финансовые метрики — включая те, что влияют на косвенные показатели, такие как удовлетворенность или скорость обслуживания, — приоритизацию рекомендуется проводить исходя из степени влияния на ключевые метрики клиентского пути (например, конверсия, время обработки запроса, уровень оттока). Это позволяет сохранить фокус на клиентоориентированности и стратегических целях, даже когда прямая финансовая оценка затруднена.

Такой подход позволяет не только корректно измерять реальный вклад ИИ, но и адекватно доносить его ценность — обеспечивая строгую финансовую аргументацию для руководства и инвесторов, а также гибкость в управлении портфелем инициатив на этапе их формирования.

Как оценить собственные цели и результаты относительно лучших рыночных практик?

Для эффективного управления цифровой трансформацией и оценки успешности внедрения искусственного интеллекта компаниям необходимо не только измерять внутренние результаты, но и сравнивать их с лучшими практиками на рынке. Это позволяет понять, где организация находится относительно других игроков отрасли и какой имеет потенциал для улучшения.

Бенчмарки — это ключевой инструмент стратегического анализа. Они служат ориентиром для оценки динамики собственных показателей относительно рынка и лидеров отрасли. Если метрики компании соответствуют или превосходят бенчмарки, значит, она движется в ногу с передовыми игроками. Если нет — это сигнал о наличии возможностей для роста и оптимизации.

В рамках методологии оценки финансового эффекта от ИИ предлагается использовать комплексный набор метрик с бенчмарками рыночными и представителей индустрии, которые позволяют провести объективное сравнение по следующим направлениям:

01 - Финансовый эффект от ИИ		02 - Операционная эффективность	
Предлагаемая метрика:	Потенциальный бенчмарк:	Предлагаемая метрика:	Потенциальный бенчмарк:
% (доля) фин. эффекта от ИИ в общем фин. эффекте компании (прибыль до налогообложения)	20% — доля финансового эффекта от ИИ в EBIT по данным консалтинговых исследований (например, McKinsey ⁸). 24% — доля финансового эффекта от ИИ в прибыли до налогообложения у Сбера ⁹ , что демонстрирует высокий уровень зрелости и масштаба применения технологий.	% оптимизированного OPEX за счет ИИ / ГЕНИИ	~10% — среднее сокращение операционных расходов (OPEX) за счет внедрения ИИ и генеративных технологий (ГЕНИИ), согласно рыночным исследованиям ¹⁰ .
Эти данные указывают на то, что представители индустрии уже получают значительную долю своих финансовых результатов за счет ИИ. Для компаний, стремящихся к лидерству, целесообразно стремиться к достижению и превышению этих показателей.		Оптимизация OPEX через ИИ — один из наиболее быстрых способов повышения рентабельности, особенно в секторах с высокой долей рутинных процессов.	
03 - Темпы роста и инвестиционная активность		04 - Уровень применения ИИ	
Предлагаемая метрика:	Потенциальный бенчмарк:	Предлагаемая метрика:	Потенциальный бенчмарк:
Среднегодовой темп роста со сдвигом на уровень зрелости	50% — среднегодовой темп роста финансового эффекта от ИИ в Сбере ¹¹ (2020–2024), что свидетельствует о ускоряющемся ROI и масштабировании решений.	% процессов / продуктов / бизнес-функций с ИИ	~80% — доля бизнес-процессов, использующих ИИ в Сбере ¹⁵ Это показывает высокую степень интеграции технологий во все аспекты деятельности.
ROI	1:8 — окупаемость инвестиций в ИИ в Сбере ¹² , что говорит о высокой эффективности принимаемых решений об инвестициях.		
Темп роста инвестиций	x1,5 — планы Сбера по росту инвестиций в ИИ до 2029 года ¹³ ; 88% — средний прирост инвестиций в ИИ в Т-Банке за 2023–2025 годы ¹⁴ .		
Эти данные подтверждают, что лидеры рынка не только активно инвестируют в ИИ, но и добиваются значительной отдачи. Компаниям важно не просто увеличивать бюджеты, но и контролировать их эффективность.			

8 McKinsey «The state of ИИ in 2023: Generative ИИ's breakout year», Август 2023
9 Результаты работы группы Сбер по МСФО за 2024 год
10 McKinsey «The state of ИИ in early 2024: Gen ИИ adoption spikes and starts to generate value», Май 2024
11 Интервью А. А. Ведяхина (<https://tass.ru/ekonomika/18713627>)
12 Интервью Г. О. Грефа (<https://m.business-gazeta.ru/news/636054>)
13 Интервью Г. О. Грефа (<https://m.business-gazeta.ru/news/636054>)
14 Внутренний прогноз по росту инвестиций в ИИ до конца 2025г
15 Выступление А. А. Ведяхина на форуме «Открытые инновации» <https://sberlabs.com/news?news=1256>

Рекомендации по применению

Слепое сравнение с рыночными бенчмарками может привести к серьезным ошибкам в оценке эффективности компании. Это происходит потому, что прямое заимствование чужих метрик без учета внутреннего контекста:

1. Игнорирует разницу в зрелости процессов

Сравнивая KPI цифрового гиганта с показателями традиционной компании, вы рискуете принять нереалистичные цели, не соответствующие текущим возможностям организации

2. Не учитывает специфику бизнес-модели

Банк с розничным фокусом и инвестиционный холдинг имеют принципиально разные метрики рентабельности и риска, что делает их прямое сопоставление бессмысленным

3. Разная методология данных

Если отраслевой бенчмарк построен на агрегированных данных по стандартизированной методике, а ваша компания использует внутренние данные с иным уровнем детализации и обработки, прямое сравнение показателей некорректно. Это подобно сравнению точности медицинского термометра и прогноза погоды — они измеряют разное, с разной целью и погрешностью.

Использование бенчмарков должно быть комплексным и учитывать специфику компании: степень зрелости процессов, систему управления данными, бизнес-модель и отраслевые особенности. Только контекстный анализ позволяет сделать объективные выводы и избежать ошибочных решений. Представленные бенчмарки носят информационно-справочный характер и не могут рассматриваться как обязательные нормативы или целевые показатели.

- Представленные метрики можно использовать как базовые ориентиры для внутреннего анализа.
- Сравнивайте свои показатели с рыночными данными (консалтинговые отчеты, публичные данные представителей индустрии) для формирования выводов о развитии ИИ в компании.
- Отслеживайте динамику изменений: например, если текущий план развития ИИ не предусматривает достижения цели в 20% финансового эффекта от ИИ в EBIT, у компании есть потенциал для расширения ИИ портфеля и увеличения эффектов от него.
- Оценивайте не только ваше текущее положение относительно рынка, но и динамику: даже при низком текущем эффекте, высокий темп роста может свидетельствовать о правильном пути развития.

Оценка финансового эффекта от ИИ должна быть не статичной, а динамической и ориентированной на сравнение с лучшими практиками. Бенчмарки по ключевым метрикам — финансовый эффект, сокращение расходов, ROI, темп роста и глубина внедрения — позволяют выстроить стратегию, направленную на достижение лидерства в цифровой трансформации. Регулярный мониторинг и корректировка подхода на основе внешних данных с учетом внутренней специфики компании помогут не только удерживать конкурентное преимущество, но и создавать долгосрочную ценность.

Какими методами можно посчитать эффект от ИИ?

Цель методики:

Оценить финансово-экономический эффект от внедрения ИИ-инициативы в бизнес-процессы организации.

Группа методов	Методы	Уровень доверия
Контролируемые эксперименты	- Вечный контроль - Разовые А/Б тесты	Высокий
Квази-эксперименты	- Propensity Score Matching (PSM) - Synthetic Control (Синтетический контроль)	Средний
Альтернативное моделирование	Сравнение текущих результатов с гипотетическим сценарием «без МЛ»	Низкий

При замере эффекта важно оценивать не просто корреляции между метриками ИИ-инициативы и бизнес-метриками компании, а именно доказать прямое воздействие ИИ-инициативы на финансовые показатели. В этом разделе описаны три группы методов для выявления причинно-следственных связей, различающиеся точностью и объемом затрат:

- Контролируемые рандомизированные эксперименты.**
Классические А/Б и глобальная контрольная группа - наиболее точные методы, считающиеся «золотым стандартом» для установления причинно-следственных связей; однако они же требуют и наибольших ресурсов.
- Квази-экспериментальные методы.**
Используются, когда А/Б тесты неприменимы, менее точны и требуют явных допущений о данных и их проверки.
- Альтернативное моделирование.**
Наименее приоритетный и точный подход; подходит для ориентировочных оценок при прозрачных предпосылках эффекта

А/Б-тесты

А/Б-тест – наиболее приоритетный и точный способ оценки экономического эффекта. А/Б тестирование представляет собой метод оценки влияния целевого воздействия на процесс, путем сравнения двух выбранных подгрупп (А – Целевая Группа – с воздействием ИИ; Б – Контрольная Группа – без воздействия ИИ).

Проведение А/Б-теста состоит из нескольких важных этапов:

1. Определение объекта воздействия, объекта рандомизации, ключевого драйвера.
2. Проверки качества сплита, формирования итоговой гипотезы, расчет минимальной достаточной выборки и выбора корректного стат. критерия.
3. Запуска теста и выставление критериев мониторинга эффекта.
4. Подведения итогов теста по выставленным критериям. Регулярный мониторинг эффекта в соответствии с выставленными критериями, и актуализация задокументированного эффекта.

☒ Объект воздействия

Объект воздействия – это тот элемент процесса, характеристики и/или количество которого может изменяться в результате внедрения ИИ модели.

Пример:

Пользователь/клиент, обращение/диалог, заказ, сотрудник/команда, счет/портфель, магазин/регион, сквозной процесс, и т.д.

Принцип выбора:

- Необходимо выбирать **наиболее верхний уровень процесса**, на котором проявляется **значимая часть изменения объекта**
- Если выбранный локальный уровень процесса не захватывает все ключевые эффекты воздействия, необходимо расширить рамки анализа **до процесса более высокого порядка** (с уровня обращения до уровня пользователя; с уровня конкретной операции до уровня бизнес-процесса, с конкретного продукта до уровня корзины продуктов и т.д.).
- При этом стоит учитывать, что чем шире определение объекта, тем сложнее будет определить воздействие. Необходимо руководствоваться принципом выбора **наименее достаточного уровня**

Примеры:

Чат-бот поддержки

Узко: обращение (время оператора, сэкономленное на одно обращение).

Шире и корректнее: пользователь — также учитываем влияние на частотность обращения, метрики удержания пользователя в канале.

Компания продаж

Узко: продукт (конверсия по продукту);

Шире и конкретнее: клиент – также учитываем конверсию по прочим продуктам корзины, частоту покупок, эластичность и т.д.

☒ Объект рандомизации

Объект рандомизации – минимальный элемент процесса с целевым ИИ-изменением, который **случайно** назначается в контрольную/целевую группы. Как правило, объект рандомизации равен объекту воздействия, однако в некоторых случаях после определения объекта воздействия требуется провести дополнительное исследование процесса, чтобы выбрать отдельный объект рандомизации.

Принцип выбора:

- Начинаем с **объекта воздействия**, выбранного в предыдущем пункте
- Проверяем на соблюдение **SUTVA**, предполагая, что объект воздействия равен объекту рандомизации
- В случае несоблюдения необходимо разделять объекты и поднимать уровень объекта рандомизации: **кластер**.

Пример:

Сотрудник → подразделение, магазин → город (все магазины в городе),
клиент → гео-зона (все клиенты в гео-зоне).

Проверяем **фиксированность назначения**: объект рандомизации и соответственно объект воздействия должен на протяжении всего А/Б теста находиться только в **одной** группе: **КГ** или **ЦГ** – не должно быть переходов из одной группы в другую.

- Критерий включения в тест: четко фиксируем объекты, которые могут попасть в эксперимент; исключаем конфликты с параллельными тестами (если возможно), а также конфликты в рамках нормативных документов, принятых в компании (например, политики коммуникации, кредитная политика или же принципы запретов на дискриминацию сотрудников).

Примечание:

SUTVA (Stable Unit Treatment Value Assumption) – допущение, что:

1. **Нет интерференции объектов:** исследуемое воздействие от ИИ на один конкретный объект не влияет на исход другой.
2. **Нет скрытых версий исследуемого воздействия:** механизм воздействия от модели ИИ одинаков для всех объектов.

Примеры необходимости отделения объектов рандомизации от объектов воздействия.

Увеличение эффективности сотрудников:

При внедрении новой системы мотивации, объект воздействия – сотрудник, однако, если разбиение на КГ и ЦГ происходит на уровне сотрудников, может случиться, что внутри одного подразделения, или, еще хуже – внутри одной команды, будут сотрудники с воздействием и без него; такая ситуация может сильно повлиять на потенциальный результат исследования, поскольку существует эффект «сарафанного радио», и такие сотрудники из КГ будут недовольны «несправедливостью», что ухудшит показатели А/Б теста и завысит дельту по метрике ключевого драйвера между ЦГ и КГ

Ценообразование:

Увеличили цены в некоторых магазинах в ЦГ, может вырасти спрос в соседних магазинах в КГ (при этом он не должен был измениться от целевого воздействия) из-за общего рынка и эластичности спроса

Алгоритм назначения сотрудников:

На часть пользователей раскатили новый алгоритм назначения курьера или сотрудника поддержки, однако совокупность сотрудников и механизм назначения общие, из-за чего изменилось время ожидания и в контрольной группе.

Неделимость:

Некоторых процессов компании: при выборе процесса, как объекта воздействия, для некоторых критических процессов невозможно подобрать случайным образом пару, чтобы разделить их на две сравнимые группы.

☑ Ключевой драйвер

Ключевой драйвер (Ключевой показатель эффективности/Ключевая метрика) — основная метрика эффекта на выбранном объекте рандомизации, по которой принимается решение о финансовой-экономической эффективности инициативы. Должна отражать **полный эффект**, а не удобный фрагмент (см. раздел принципы финансовой оценки).

Принципы выбора:

- **Полнота экономики.**

Необходимо учитывать все инкрементальные **доходы** и **затраты**, на которые потенциально может повлиять модель.

Пример:

Модель оптимизирует продажи продукта А для всех клиентов, однако в то же время можем повлиять на продажи продукта Б, С, Д и т.д. – тогда необходимо так же учитывать эффект во всей линейке продуктов (каннибализация и кросс-продажи). Ключевыми драйверами в такой ситуации должны быть выбраны продуктовые метрики всех связанных продуктов полки продаж.

- **Единица расчета.**

На формулу расчета ключевой метрики не должно оказывать влияние исследуемое изменение от внедрения ИИ. Метрику необходимо рассчитывать на фиксированной популяции объектов, которые участвуют в тесте, не смещая и не искажая расчет. Воздействие в целевой группе часто меняет число событий (диалогов, показов, доставок и т.д.), поэтому метрики, взятые в отношении «на событие», становятся зависимыми от воздействия. Правильный подход — считать на **единицу** объекта рандомизации.

Пример:

Можем считать конверсию по показам, однако ИИ модель в том числе может влиять на кол-во просмотров, что может исказить расчет, в таком случае лучше всего считать клики на пользователя за период.

- **Окно агрегирования.**

Фиксируем **измерительное окно ключевой метрики** (7/30/90 дней и т.д.), где материализуется значимая часть эффекта. Необходимо выбирать **минимально достаточное окно** по задержки эффекта во времени и мощности теста.

- Направление и пороги.

Четко задаем направление улучшения метрики: «выше лучше/ниже лучше»; заранее фиксируем **MDE** (минимально полезный эффект) и бизнес-порог принятия решения.

Простые формулы вместо «индексов».

По возможности необходимо использовать разницу средних базовых величин (деньги/время/счетчики), избегать доли от долей и сложных агрегаций.

Борьба с манипуляциями и защитные метрики.

Ключевая метрика не должна стимулировать вредные практики, так же необходимо добавить 2-5 защитные метрики.

- *Борьба с манипуляциями подразумевает совокупность мер, направленных на борьбу с искажениями и манипуляциями данными. Включает в себя фиксацию групп и метрик на протяжении всего теста, фиксацию протокола проведения теста и невозможность менять «правила игры» по ходу А/Б теста.*
- *Защитные метрики – это метрики, измеряющие потенциальные потери, которые может понести компания в ходе тестирования воздействия от ИИ. Необходимо четко прописать алгоритм принятия решения о досрочной остановке А/Б теста и пороги значимости потерь, для случая, если тест будет наносит вред процессу/финансовому состоянию/клиентам компании.*

Проверка качества сплита

После корректного **выбора объекта воздействия, единицы рандомизации и ключевого драйвера** необходимо убедиться, что сплит на группы выполнен без систематических перекосов. В противном случае выводы эксперимента не **будут достоверными**, что сделает невозможным их тиражирование и практическое применение для получения финансового эффекта.

Что необходимо проверять:

- Несоответствие долей выборки (SRM)

Проверка соответствия фактических долей контрольной и целевой групп плановому теоретическому распределению. Проверку необходимо проводить как по всем тестовым выборкам, так и по ключевым срезам (платформа, версия, регион).

Пример:

Если тест был спланирован с предполагаемым распределением 50/50, а после запуска размеры групп статистически значимо отличаются (например, 60/40), это явный признак некорректного проведения эксперимента.

- **Баланс ковариат.**

Проверка на сопоставимость контрольной и тестевой групп по заранее заданному набору признаков (признаков однородности объекта рандомизации): возраст, регион, платформа, пре-метрика и т.д.

Важно: в качестве признаков нельзя брать какие-либо пост-тритмент метрики, рассчитанные за период после запуска теста, которые могут теоретически измениться после применения воздействия. В качестве признаков необходимо брать свойства/метрики, которые **известны, что до запуска и не меняется из-за воздействия**: прошлое поведение клиента (покупки/клики/обращения за n дней до теста), стабильные свойства объекта рандомизации (регион, тип клиента, устройство/платформа, тариф) и контекст попадания в тест (канал, день недели/время суток и т.д.). Для определения всех релевантных признаков однородности, требуется изучить бизнес-процесс, в который внедряется ИИ, а также объект рандомизации, который был выбран для проведения А/Б теста.

- **А/А тест.**

Он также должен выступать инструментом проверки качества сплита и сравнимости групп тестирования. А/А тест подразумевает собой полный аналог А/Б теста, но без целевого воздействия в одной из групп. Как правило, используется А/А тест для проверки однородности собранных групп в том числе и по ключевой метрике (ключевому драйверу) к моменту старта А/Б теста, перед запуском А/Б теста.

- **Фиксированность назначения** – не должно быть объектов рандомизации, которые находятся одновременно в контрольной и целевой группах во время теста, или могут переходить из одной группы в другую по ходу теста – **нет переназначений**

Что необходимо проверять:

1. **Перед запуском** (предварительный запуск на 1–5% трафика) — убедиться, что разделение по группам выполнено корректно.
2. **Каждый день в ходе теста** — автоматизированный мониторинг показателей по всем заданным срезам данных с системой оповещений.

Финальная гипотеза, стат. критерий и размер выборки

Далее необходимо сформулировать итоговую проверяемую гипотезу и выбрать стат. критерий.

Гипотеза в рамках А/Б теста формируется в терминах дельты (инкремента) ключевой метрики замера эффекта относительно выбранного бейзлайна. В качестве бейзлайна должна быть выбрана лучшая ближайшая альтернативу процессу TO BE.

Гипотеза по ключевому драйверу/ключевой метрике формулируется в формате:
 $H_0: \Delta = 0$, $H_1: \Delta \neq 0$

По умолчанию используем двустороннюю проверку гипотезы (дельта эффекта может быть как положительной, так и отрицательной). Односторонняя проверка может быть использована, если заранее доказано, что обратный эффект невозможен.

Рекомендации по стат. критериям:

1. Для непрерывных метрик по умолчанию используется Welch t-test (необходима независимость наблюдений)
2. Для доли/бинарных распределений – z/χ^2 критерий.
3. Для ненормальных распределений возможно использовать ранговые критерии и методы, например U-test (однако стоит учитывать специфику проверяемой гипотезы данным критерием)
4. В случае метрик-отношения: дополнительно применяется Delta-метод или Линеаризация
5. В случае нестандартных агрегаций (медиана, квантиль) или неприменимости критериев, описанных выше, возможно применение **Bootstrap**

Далее формулируем статистическую (далее – стат.) значимость и мощность эксперимента. Базовая рекомендация – **уровень стат. значимости 5%, мощность 80%**. Данные параметры могут быть изменены основываясь на толерантности компании к ошибкам I и II рода, однако не рекомендуется выбирать менее строгие значения параметров.

Если гипотез в рамках одного теста несколько, то необходимо делать поправки на множественное сравнение. В зависимости от конкретного случая, это может быть сделано, например, поправкой Бонферрони/методом Бонферрони-Холма/и т.д.

Размер выборки и длительность эксперимента

Основываясь на параметрах ключевого драйвера, которые были определены ранее необходимо рассчитать размер минимальной достаточной выборки и длительности теста.

Размер выборки будет в первую очередь зависеть от выбранного уровня стат. значимости, мощности, MDE и дисперсии ключевой метрики.

Формула расчета размера выборки в базовом кейсе:

$$N \approx \frac{(z_{1-\alpha/2} + z_{\text{power}})^2}{\text{MDE}^2} \times \left(\frac{\sigma_1^2}{s} + \frac{\sigma_2^2}{1-s} \right)$$

При заданных параметрах $\alpha=5\%$, $\text{power} = 80\% \rightarrow \approx 1.96 \approx 0.84$
 s - доля целевой группы в множестве (КГ+ЦГ), σ_1^2 – дисперсия ЦГ, σ_2^2 – дисперсия КГ.

По вышеописанной формуле определяется количество наблюдений, требуемое, чтобы доказать достижение значения MDE по ключевой метрике, и что найденная дельта ключевой метрики действительно вызвано ИИ воздействием, а не является «шумом» наблюдения в рамках дисперсии.

Таким образом, $N_{\text{цг}} = sN$ и $N_{\text{кг}} = N - N_{\text{цг}}$. Зная число наблюдений, которое можно собрать в один месяц, можно рассчитать требуемое число месяцев для подтверждения эффекта на А/Б тесте.

Важно также отметить, что точная формула расчета N может отличаться в зависимости от выбранного дизайна и критерия теста.

Далее необходимо определить длительность теста на основе размера трафика, требуемого минимального размера выборки, а также срока вызревания ключевой метрики, если срок вызревания отличается от минимального требуемого срока, рассчитанного по формуле. При определении длительности также стоит учитывать эффект новизны и сезонность. Например, необходимо запускать тест минимум на одну полную неделю, если есть недельные паттерны в данных, даже если необходимый размер выборки набирается за меньшее количество дней. Рекомендацией здесь может служить следующее: для пилотирования ИИ-инициативы и проведения А/Б теста необходимо брать срок, полученный через расчет MDE + срок вызревания эффекта (окно агрегирования).

☑ Ускорение тестов

Возможна ситуация, когда на этапе дизайна требуемый срок проведения оказывается неоправданно большим. В таком случае процесс оценки нарушает принцип материальности: затраты на проведения пилота оказываются сопоставимы, или, в крайних случаях, даже больше потенциального дохода от самой ИИ модели.

В данном случае стоит обратиться к методам ускорения А/Б-тестов:

Стратификация

Стратификация — это метод снижения дисперсии. Его суть в том, что популяцию **заранее разбивают** на однородные группы (**страты**) по ключевым доэкспериментальным признакам, а затем проводят рандомизацию внутри каждой группы. При анализе результаты **взвешенно объединяются** с учетом размера каждой страты. Такой принцип «**сравнения похожего с похожим**» позволяет исключить из оценки межгрупповую дисперсию, оставив лишь дисперсию внутри групп. В **результате** требуемый объем выборки или длительность теста уменьшается.

Принцип работы:

1. Необходимо выбрать 1-3 сильных признака объекта рандомизации независимых от эффекта воздействия (например, регион или платформа)
2. Задать веса равным долям трафика на пред-тестовых данных
3. Рандомизировать объекты внутри каждой страты по плановой доле КГ/ЦГ
4. При анализе использовать взвешенную метрику

$$\hat{Y}_{\text{strat}} = \sum_{k=1}^K \pi_k \bar{Y}_k, \quad \pi_k = \frac{N_k}{N}, \quad \text{где } \bar{Y}_k - \text{среднее в страте } K$$

Когда применять.

Внутри популяции объектов воздействия есть четкие группы с отличающейся ключевой метрикой (ОС, страны/регионы, каналы). Чем больше различия между стратами и стабильнее распределение трафика, тем выше польза.

Важно:

1. Не нужно чрезмерно дробить трафик, необходимо держать разумный размер страты, редкие категории объединять в «other»
2. Страты необходимо фиксировать до запуска и проверять SRM по ним

CUPED (ковариатная коррекция доэкспериментальной метрикой)

CUPED - метод снижения дисперсии за счет вычитания из итоговой метрики части дисперсии, объясняемой пред-тестовой метрикой (значением метрики до старта теста или другой инвариантной к тестовому воздействию величиной).

Принцип работы:

Берем итоговую метрику Y и пре-метрику X на том же объекте рандомизации (пользователь/клиент/зона), считаем преобразованную метрику

$$Y^* = Y - \theta(x - \bar{x}), \theta = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}$$

Далее сравниваем Y^* между тестом и контролем обычным способом (Welch t-test). Дисперсия новой преобразованной метрики $\text{Var}(Y^*) = \text{Var}(Y)(1-\rho^2)$, где $\rho = \text{corr}(Y, X)$.

Соответственно, чем больше корреляция ковариаты и изначальной метрики, тем эффективнее будет работать метод.

Когда применять

Есть явная стабильность поведения объектов воздействия во времени: прошлые покупки, частота обращений, средний чек и т.п.

Важно:

X должна быть строго инвариантна к эффекту воздействия теста. В идеале это должна быть метрика, рассчитанная до попадания в тест.

Глобальная Контрольная Группа (Holdout)

Глобальная контрольная группа (ГКГ) — это постоянная группа, формируемая из подходящей популяции объектов, которая не получает новых ML-моделей (либо получает только стабильную базовую версию). ГКГ помогает отслеживать долгосрочный дрейф данных и поведения объектов, а также выявлять потенциальную

деградацию модели. Ее использование позволяет непрерывно контролировать эффект в рабочей среде и более корректно оценивать вклад множества параллельных экспериментов. ГКГ по своей сути является непрерывным А/Б-тестом. Метод стоит рассматривать как дополнение к разовым А/Б-тестам, а не как замену. ГКГ позволяет более точно и честно оценить интегральный экономический эффект, но требует значимых усилий в реализации и поддержке.

При решении о создании ГКГ стоит руководствоваться принципом материальности, в большинстве случаев затраты на оценку могут превышать потенциальную пользу. Рекомендуется использовать ГКГ для **крупных ML-продуктов**, которые приносят значимую долю от общего экономического эффекта от ИИ в компании и/или для ML-продуктов с выделенной командой, которая на постоянной основе улучшает модель.

Дизайн:

Состав.

В ГКГ входят те же объекты, удовлетворяющие условию включения, что и в разовый А/Б-тест.

Размер.

Обычно 1–5% трафика, стоит ориентироваться на тот же MDE, что в разовом А/Б. Доля фиксируется заранее.

Фиксированность назначения.

Один и тот же объект всегда остается в ГКГ (если не было намеренного перемешивания), не должно быть переназначений.

Логика сплита.

Та же, что и в разовом А/Б-тесте. Важно, чтобы новые пользователи (объекты) тоже добавлялись в ГКГ согласно заранее выбранной доле.

Изоляция.

ГКГ исключается из всех экспериментов.

Метрики.

Те же ключевые и защитные метрики, что и в обычном А/Б-тесте.

Анализ.

В качестве драйвера экономического эффекта считается дельта ключевой метрики, усредненная по объектам рандомизации в окне W (например, 7/30 дней). Необходимо смотреть оценку и 95% ДИ.

Частые проблемы:

1. Нестабильность отбора.

Главной проблемой ГКГ является сложность технической реализации и поддержание репрезентативной ГКГ на долгом горизонте. Репрезентативность легко ломают разовые сбои сплит-алгоритма или невнимательная реализация новой функциональности/модели. Типичные причины: нестабильные идентификаторы (слияние/разделение), незаметные изменения критериев включения в тест, коллизии с параллельными экспериментами.

2. Нерепрезентативность ГКГ.

Слишком маленькая доля ГКГ дает высокий шум и «прыгающие» оценки, перекося по сегментам (платформа, регион, тип клиента, активность) делает непереносимой на весь портфель, со временем дрейф трафика/изменения критериев включения может приводить к тому, что группа перестает описывать текущую популяцию.

Меры по предотвращению проблем:

- Стабильный идентификатор объекта рандомизации
- Единый источник флага принадлежности ГКГ
- Запрет переназначений между группами
- Четкая фиксация критериев попадания в ГКГ
- Полное перемешивание ГКГ при любых изменениях критериев
- Обязательные мониторинги и оповещения по здоровью ГКГ: переназначения групп, контроль слияния/разделения аккаунтов, SRM по всей популяции и по ключевым срезам.

Пересборка Глобальной Контрольной Группы:

При обнаружении перекося в ГКГ необходимо остановить анализ, зафиксировать, устранить причину и выполнить перемешивание ГКГ. Создать новую версию ГКГ параллельно текущей, запустить в теновом режиме (без влияния на объекты воздействия) на 3–7 дней, проверить равномерность сплита по всему трафику и по ключевым срезам и проверить баланс ковариат. Переключить продукт на новую версию ГКГ и выдержать буферный период, данные переходного окна в оценку не включать (длина буферного периода должна быть не менее длины окна замера эффекта W). Далее считать экономический эффект по новой версии ГКГ.

Switchback-тестирование

Принцип работы

Воздействие «тест/контроль» применяется не по объектам, а по временным слотам, регулярно «переключая» назначение. Подходит для систем с сильной интерференцией: ценообразование, аллокация операторов/курьеров, ранжирование, где общий рынок/совокупность ресурсов «перетекает» между группами.

Пример:

Динамическое ценообразование курьеров: переключаемся по часовым блокам внутри города и сравниваем выручку на пользователя и время доставки, контролируя «час×день».

Ключевые допущения

1. Временные слоты сравнимы после учета сезонности/шоков, перенос эффекта между слотами либо отсутствует, либо гасится буферным окном.
2. Назначение слотов в «тест/контроль» действительно случайно (или псевдослучайно) и сбалансировано по календарным факторам.

Дизайн

Единица рандомизации.

Временной слот (например, 30–60 мин) × локация/кластер.

Метрики.

Те же ключевая и защитные метрики, агрегированные в окне W.

Расписание.

Случайные или чередующиеся блоки Тест/Контроль с равными экспозициями по часам/дням. Необходимо добавлять период вымывания эффектов при смене состояния, который не будет учитываться в оценке.

Риски

1. Неверная длина слота.
Слишком короткий: шум/перетекания, слишком длинный: низкая мощность.
2. Непокрытая сезонность.
Попадание в «сильные» часы только тестом.
3. Переток эффекта в соседние слоты.
Лечится буферным окном.

Квази-экспериментальные методы

Если провести рандомизированный эксперимент невозможно — из-за юридических или технических ограничений, высокой стоимости или нарушения условий его применимости (SUTVA), — можно использовать альтернативные квази-экспериментальные методы. Однако стоит учитывать, что по силе причинной идентификации эффекта они значительно уступают классическим А/Б и опираются на довольно жесткие допущения о данных. В данной секции описаны только базовые принципы работы квази-экспериментальных методов, перед практическим применением стоит отдельно внимательно изучить тонкости и детали использования методов, а также учесть ограниченную область их применимости.

Propensity Score Matching (PSM)

Принцип работы

Метод позволяет сравнить тех, на кого было оказано воздействие, с очень похожими

объектами, среди тех, на кого воздействия не было, при раскатке изменения сразу на всю популяцию объектов воздействия, без проведения честного А/Б-теста. С помощью ML-модели подбираются «близнецы» по наблюдаемым признакам: строится модель предсказывающая вероятность попасть под воздействие новой функциональности/алгоритма (propensity score). Далее объекты с фактически примененным воздействием и без сопоставляются по значению propensity score, тем самым группы выравниваются по наблюдаемым факторам. После сравниваются метрики двух групп, средняя разница и есть оценка эффекта для получивших воздействие. Метод подходит для случаев, когда рандомизация невозможна, но есть сильные пред-тестовые данные, которые описывают различия между группами, получивших и не получивших воздействие.

Пример:

Запуск ИИ-ассистента продаж для B2B-клиентов: ассистент раскатывается на всех клиентов, затем подключившиеся клиенты матчатся с теми, кто не пользовался ассистентом по отрасли, размеру, прошлой выручке, и тд. Считается дельта средней выручки по сматченным группам.

Ключевые допущения

1. При условии фиксированных ковариат назначение в тест «как бы случайно».
2. Есть перекрытие распределений propensity score между группами

Дизайн

Модель.

Логистическая регрессия или бустинг

Проверки.

Баланс ковариат

Ковариаты.

Только пред-тестовые данные (история покупок/обращений, чек, гео, устройство, тариф и т.д.).

Риски

1. Ненаблюдаемые сильные ковариаты.

Пример:

В пилот LLM-помощника взяли только звезд-операторов по личной оценке руководителя, о чем не знала продуктовая команда.

2. Плохое перекрытие распределений.

Среди получивших воздействие нет похожих объектов на не получивших.

Пример:

Все получившие воздействие пользователи iOS, не получившие преимущественно пользователи Android.

Синтетический контроль (Synthetic Control)

Принцип работы

Метод позволяет оценить причинный эффект интервенции для одной или нескольких агрегированных единиц (город, линия бизнеса), когда классический А/Б невозможен. Для объекта, к которому применено воздействие, строится синтетический двойник из нескольких объектов, которые не получили воздействие. Подбирается взвешенная смесь донорских объектов, которая очень точно повторяет поведение объекта с примененным воздействием на пред-тестовых данных. Разница ключевой метрики между оригинальным объектом и синтетическим близнецом на периоде после интервенции является оценкой эффекта воздействия.

Пример:

В райдтех сервисе в Казани включили новый алгоритм распределения курьеров, а в похожих городах (Нижний Новгород, Самара, Уфа) оставили старый алгоритм. Из похожих городов собрали синтетического двойника как $0.3 \text{ Новгорода} + 0.5 \text{ Самары} + 0.2 \text{ Уфы}$, который довольно точно повторял траекторию Казани до запуска изменения.

Ключевые допущения

1. Аппроксимация: объект с воздействием в пре-периоде хорошо описывается выпуклой комбинацией объектов-доноров.
2. Отсутствие интерференции между единицами: воздействие не искажает поведение объектов-доноров.

Риски

1. Плохая подгонка на пре-периоде:
Синтетический близнец не похож на факт
2. Слишком короткая история наблюдений перед тестом

3. Сильный внешний шок на пост-периоде.
Он оказывает различное влияние на объекты. Например, новый закон или COVID.

Альтернативное моделирование

Приблизительная, экспертная оценка экономического эффекта на основе бизнес-моделирования. Строится модель того, как система должна работать, и рассчитывается порядок величины эффекта при заданных допущениях. Не является квази-экспериментом и доказательство причинного влияния.

Когда применяется.

1. Стоит применять только в крайнем случае: рандомизация и квази-эксперименты невозможны (правовые/этические ограничения, сильная интерференция, редкие события, очень малые выборки)
2. ML-инициатива потенциально не приносит значимую часть от общего эффекта вклада ИИ в компании

Пример построения

- Базовый сценарий:
Прогноз без ИИ (сезонность, тренды).
- Механика эффекта:
Явная цепочка «применение ML → изменение метрики → деньги».
- Сценарии и неопределенность:
Base/Low/High + при необходимости Монте-Карло симуляция по ключевым неопределенностям.
- Юнит-экономика и ограничения:
Выручка/капитальные расходы/операционные расходы, комиссии, операционные бутылочные горлышки.
- Валидация здравым смыслом:
Бенчмарки по прошлым запускам/рынку, консистентность с PnL.

Предлагаемые решения сложных вопросов на базе best practices лидеров индустрии

Как измерять эффект, где нет денежных метрик? Косвенные стратегические метрики

Ключевые стратегические нефинансовые эффекты от внедрения ИИ могут выражаться как в метриках конкретного продукта, в который внедряется ИИ (например, NPS, CSI, retention), так и в экосистемных метриках (например, привлечение в экосистему, retention в экосистеме, лояльность клиентов).

Самый простой классический пример экосистемного эффекта в банковской сфере:

Прирост в выдачах дебетовых карт может иметь околонулевой эффект на PnL текущего года, однако выдача дебетовой карты может повышать вероятность выдачи кредитных продуктов в будущем, что может принести значимый положительный финансовый эффект.

В крупных экосистемах могут присутствовать гораздо менее очевидные и менее линейные скрытые эффекты.

Ключевое в этом вопросе - это умение находить прокси-метрики на данных в настоящем, которые могут прогнозировать активность клиентов в будущем. Подобные экосистемные метрики могут быть переведены в PnL текущего периода с помощью исследовательских А/Б-тестов, направленных на замер размена экосистемных метрик с деньгами.

Пример:

Метрика лояльности – вовлеченность клиентов в экосистему.

1. С помощью А/Б теста измеряем курс размена пункта лояльности с денежными метриками и получаем курс размена, например, через применение прямых денежных стимулов: доп. кэшбэки/скидки (или наоборот уменьшение скидок и кэшбэков);
2. Через А/Б тест измеряем, как новый ИИ-продукт влияет на лояльность и PnL, получаем некий курс лояльности X. Важно помнить, что при этом курс размена не будет являться универсальным и нельзя транслировать его на другие кейсы;
3. Сравниваем курс лояльности X в ИИ-продукте с курсом размена Y (из п.1).

Если с помощью ИИ-продукта мы можем «купить» лояльность дешевле, чем через тривиальные изменения денежных стимулов, то ИИ положительно влияет на PnL. Но есть инициативы, где крайне сложно, а порой и невозможно перевести метрику в финансовый эффект, но это не значит, что такие метрики не стоит учитывать при выборе ИИ проекта.

Мы предлагаем 3 группы потенциальных стратегических эффектов от ИИ с примерами метрик:

Удовлетворенность / репутация

- Рост лояльности
- % персонализации коммуникаций (от всех коммуникаций)
- Стоимость бренда / медиа охват с учетом сентимента

Эффективность

- Объем бизнеса на 1 сотрудника (например, доходы с КСБ на 1 сотрудника КИБ КСБ¹⁴ или кол-во закрытых обращений в КЦ¹⁵ на 1 сотрудника КЦ)
- % автономизации критических процессов (например, доля выданных кредитов через ИИ по сегментам клиентов)
- Охват ролей / сотрудников ИИ-помощниками

Удовлетворенность / репутация

- Loss prevention
- % детекций фрода
- Объем утечек данных (кол-во утечек X кол-во затронутых клиентов)

Упомянутые метрики необходимо рассматривать индивидуально под каждый конкретный проект и, возможно, адаптировать их с учетом специфики и целей. Эти показатели не являются универсальными или обязательными для использования, но они основаны на многолетнем опыте оценки проектов представителей финансовой отрасли. Поэтому подход к выбору и применению метрик должен быть гибким и ориентированным на достижение максимальной ценности для бизнеса в каждом отдельном случае.

Как определить тип и объем инвестиций в R&D?

“Фундаментальный” R&D - развитие науки и поддержка научной активности, в т.ч. развитие Practical AGI

14 КИБ КСБ - Корпоративно-инвестиционный бизнес, сегмент Крупный и средний бизнес

15 КЦ - Колл-центр

“Прикладной” R&D - развитие бизнеса, требующее инвестиций на данный момент, которое потенциально принесет денежный эффект через некоторое время, имеет бизнес-заказчика. В рамках “Прикладного” R&D можно отдельно выделить платформенный ИИ, который создается специально для владельцев платформ.

Тип R&D	Определение/критерии	Как учитываем в эффекте от ИИ
“Прикладной”	- Основная метрика – конверсия в прод - Наличие бизнес-заказчика, техническая реализуемость, научная новизна - Наличие прямого или косвенного бизнес-эффекта / влияния на метрики бизнеса, включая платформенные метрики (в т.ч. GR) - Горизонт планирования 1-3 лет	Показываем кумулятивно расходы в PnL по определенному R&D
“Фундаментальный”	- Основная метрика – научная публикация - Фундаментальные исследования - Отсутствие определенного бизнес-результата по итогам исследования - Горизонт планирования до 10 лет	Отдельная статья, не включаем расходы в эффект от ИИ

Как распределить бюджет на R&D?

Как один из интересных примеров распределения бюджета на R&D, можно привести пример опыта компании Huawei¹⁶. Huawei ежегодно инвестирует около 180 миллиардов юаней в исследования и разработки, при этом важное разделение средств компании отражает разные цели и сроки исследований. Около 60 миллиардов юаней направляется на фундаментальные исследования — это долгосрочные проекты, результаты которых не поддаются немедленной оценке по критериям эффективности или коммерческой выгоды. Эти исследования создают базу для технологических прорывов и новых знаний, которые могут проявиться только через годы или даже десятилетия, но без них невозможен устойчивый инновационный рост. Еще 120 миллиардов юаней тратятся на прикладные исследования и разработку продуктов, где эффективность и результаты оцениваются по конкретным рыночным и коммерческим показателям. Это позволяет компании реализовывать новые технологии на практике,

16 Интервью Жэнь Чжэнфэя из Huawei (国家越开放，会促使我们更加进步)

создавать востребованные продукты и сохранять конкурентоспособность на быстро меняющемся рынке. Такое четкое разделение и баланс между фундаментальным и прикладным R&D позволяют Huawei не только строить инновации с опорой на глубокие научные открытия, но и оперативно реагировать на требования промышленности и бизнеса, что особенно важно в условиях текущих глобальных вызовов.

Как измерять эффект от R&D?

Цель R&D	Наличие бизнес-эффекта	Предлагаемая оценка	Возможные KPI
“Фундаментальный” R&D	Отсутствует	Ценность проекта возможно оценить внешним сообществом ученых	- Кол-во и статус публикаций (A*) - Цитируемость Инновационность (оценивается учеными)
“Прикладной” R&D	Присутствует	- Ценность проекта оценивается в момент перехода проекта в продукт и получения финансового эффекта через его продажи - При этом у команды могут быть не только успешные проекты - Эффект от успешных проектов должен «перекрывать» неудавшиеся запуски	- Процент конверсии в успешные проекты/ доля успешных проектов - Кол-во продаж успешного продукта

Важно:

При расчете эффекта от R&D расходы учитываются кумулятивно по технологии, а не отдельно по каждому проекту.

Основные метрики R&D – конверсия в прод и научные публикации.

Как избежать двойного подсчета эффекта от проектов?

Определение объектов двойного учета эффектов

Гранулярность расчета эффектов

В зависимости от поставленной цели, финансовый эффект от изменения в бизнес-процессе может рассчитываться на различных уровнях – от наибольшего макроуровня влияния на весь процесс до микроуровня влияния на конкретный компонент процесса.

Объектом финансовой оценки является внесение изменения в операционную модель бизнес-процесса (стр.6). Преимущественно используются следующие уровни расчета эффектов:

- Инициатива
- Продукт/канал/отдельная ветка бизнес-процесса или технологии
- Модель

ИИ-инициатива предполагает комплексное изменение бизнес-процесса, предполагающее внедрение/модификацию моделей, изменение критериев и настроек системы принятия решений. Как правило, ИИ-инициатива предполагает серьезное изменение бизнес-процесса.

Пример ИИ-инициативы:

Внедрение «умного ценообразования» в механизм кредитного конвейера. Данный подход предполагает внесение существенных изменений в логику назначения оптимальной ставки кредитования и встраивание модельной логики в кредитный процесс.

ИИ-инициатива может быть как частью одного продукта, так и предполагать внесение изменений в компоненты нескольких продуктов, каналов и прочих модальностей/технологических процессов. В то же время, ИИ-инициатива может быть частью одного продукта, но предполагать разработку группы моделей.

Выбор уровня гранулярности для расчета финансового эффекта зависит от специфики процесса, переводимого на модельный подход. Важно учитывать, что в настоящее время модельный ландшафт многих финансовых организаций насчитывает сотни моделей машинного обучения и искусственного интеллекта. В этой связи расчет

финансового эффекта на уровне каждой отдельно взятой модели возможен и/или финансово целесообразен далеко не всегда. К расчету эффекта на уровне модели крупные организации прибегают в случаях, когда инициатива по изменению процесса состоит из одной модели или же когда модель имеет высокий уровень значимости для бизнеса.

Классы потенциальных задвоений

Финансовыми организациями принято выделять пять наиболее характерных причин потенциальных задвоений финансовых эффектов:

1. Технологические и архитектурные причины:

Модульная многоступенчатая логика применения моделей в бизнес-процессе (модульные/каскадные модели) и сложность ее учета в оценке финансового эффекта. Каскадные модели – это группы моделей, направленные на один и тот же объект управления (продукт, услугу, клиента).

Пример:

Модель скоринга, построенная на нейросетевых признаках, или модели вероятности дефолта, представляющая собой ансамбль из нескольких модулей (транзакционный модуль, групповая поддержка, предупреждающие сигналы, пр.)

2. Процессные причины:

Реализация параллельных бизнес-инициатив несколькими командами с общим КПЭ.

Пример:

Параллельные склонностные модели для разных каналов продаж с общей метрикой «конверсия», каждой из команд независимо может быть оценен финансовый эффект, который отражает степень влияния моделей на один и тот же процесс

3. Методологические причины:

Расчёт комбинированных, косвенных и взаимозависимых метрик. Комбинированная метрика – это метрика, зависящая от нескольких факторов.

Пример:

LTV, где часть эффекта зависит от оттока; модель кредитного скоринга влияет

на одобрение заявок, что, в свою очередь, меняет нагрузку на контакт-центр. Таким образом, выбор неверной метрики может отражать финансовый эффект не только от модели, но и от всего бизнес-процесса.

4. Операционные причины

Повторное применение одной модели/сервиса в разных процессах или каналах.

Пример:

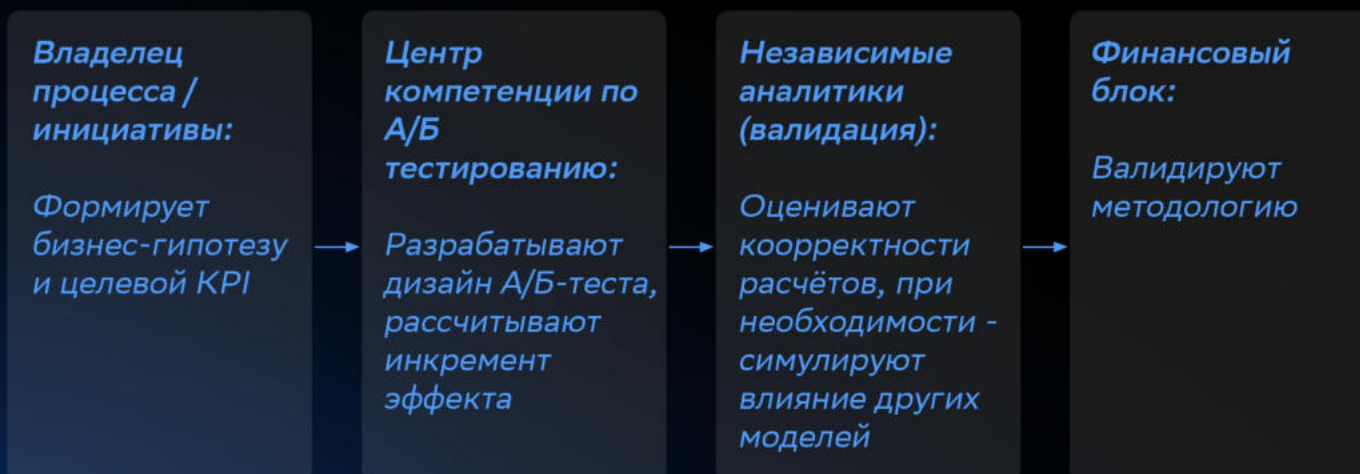
Применение модели дохода клиента для процесса кредитования и для процесса удержания. При этом важно учитывать взаимозависимость эффектов в различных процессах, в которых применяется каждая конкретная модель.

5. Расчет эффекта от новой версии модели в текущем процессе без учета инкрементальности эффекта

Эффектом новой версии модели признается не очередной инкремент, который она вносит в процесс, а общий эффект по сравнению с отсутствием модели в целом, что приводит к существенному завышению результатов. (детализировано на стр.22).

Процесс расчёта финансового эффекта

Для получения наиболее объективного результата процесс расчета финансового эффекта на всех этапах жизненного цикла модели от формирования гипотезы до финального подтверждения в расчет вовлекаются несколько независимых подразделений. Ключевые участники процесса описаны в главе «Зачем и кто оценивает фин. эффект в рамках ИИ трансформации?» (стр.2).



Если ИИ-инициатива имеет относительно небольшой фин. эффект или является нематериальной, то допустимо применение упрощенного процесса оценки (с сокращенной валидацией и выборочными проверками).

Способы ведения системного учета влияния ИИ-моделей на бизнес-процессы банка

Системный учет влияния ИИ-моделей на бизнес-процессы – инструмент, позволяющий повысить качество управления модельным риском, в том числе в части контроля ситуаций возникновения задвоения расчета финансового эффекта.

Элементами системного учета являются:

- Формализованный и утвержденный на уровне всей организации подход и методика оценки фин. эффектов ИИ с закрепленным распределением ролей участвующих в оценке подразделений.
- Реестр бизнес-процессов компании с маркировкой подпроцессов, переведенных на модельный подход, едиными уникальными идентификаторами ИИ-моделей, работающих в процессах.
- Реестр финансовых, продуктовых и математических метрик на принятом в компании уровне гранулярности расчета эффектов с маркировкой потенциального модельного риска; предполагается учет как потенциальных метрик, рассчитанных при возникновении идеи о реализации инициативы, так и фактических, полученных на стадиях от прототипирования до внедрения и регулярной валидации.
- Реестр бизнес-документации по ИИ-инициативам.
- Реестр технической документации по ИИ-моделям.
- Иные артефакты по моделям ИИ, являющиеся обязательными с точки зрения внутренних регламентных процедур организации.

Методика оценки двойного эффекта

Инструменты разделения вклада ИИ-инициатив и избегания двойного учета

1. Для количественного разделения вклада инициатив на уровне моделей применяются методы множественного А/Б тестирования, как последовательного, так и параллельного. Процесс А/Б тестирования подробно описан в главе «Какими методами можно посчитать эффект от ИИ» (стр. 32).
- В случаях, когда несколько моделей влияют на одного клиента / сотрудника – используется техника разведения пилотов во времени для того, чтобы изолировать вклад каждой модели.
 - В случаях, когда одна модель используется в нескольких процессах, рассчитывается эффект от применения модели в каждом процессе на уровне гранулярности, принятом в компании.
 - Если же требуется оценить эффект от «типа модели», например, ИИ-модель дохода

клиента используется как для кредитного скоринга, так и для процессов антифрода, то эффект рассчитывается как сумма эффектов на уровне модели.

2. Для операционного разделения вклада инициатив используются инструменты ведения системного учета (п. 1.4), в том числе присвоение ИИ-моделям уникальных финансовых идентификаторов с занесением в реестры бизнес-процессов и метрик.
3. В сложных случаях, в том числе при предпосылках канибализации или отсутствии экономической целесообразности в проведении множественного тестирования, используются:
 - Процедуры укрупнения расчетов на уровень выше, к примеру, с уровня модели до каскада моделей, с уровня инициативы до уровня продукта и т.д.
 - Процедуры приоритизации, к примеру, эффект может засчитываться основной самой доходной модели в инициативе / продукте / каскаде.

Примеры:

Для сложных процессов, таких как кредитование, где каждый этап процесса покрыт моделями, рассчитываются эффекты в двух разрезах: эффекты по каждой из моделей (склонности, PD, ценообразование и т.д.) и общий эффект от всех ИИ внедрений на процесс выдачи кредита. Это позволяет учесть пересечения эффектов и правильно оценить суммарный прирост.

Для сложных моделей, где есть и бустинги и нейросети, рассчитывается эффект от каждого компонента, а также суммарно от всей сложносочиненной модели.

При изменении версии текущей модели в эффект учитывается только инкремент по сравнению с предыдущей версией, что позволяет корректно оценить влияние дополнительных трудозатрат в привязке к дополнительной ценности.

4. Для контроля учета используется независимая экспертная оценка проводимых расчетов с целью выявления потенциальных процессов и зон, подверженных двойному учету финансовых эффектов от внедрения моделей.

В каких случаях не учитывается эффект от отдельно взятых моделей

Существует четыре типа ситуаций, в которых расчет финансового эффекта не производится:

1. Если эффект незначительный

2. Если недостаточно данных для корректной оценки вклада инициативы

3. Если модель не дает самостоятельного результата (в таком случае используются методы укрупнения, описанные в п. 2.1)

4. Если самостоятельный результат отдельно взятой модели не представляет интерес для бизнеса

Примеры риска возникновения двойных эффектов

Процессы, наиболее подверженные риску двойного учета

В финансовых организациях наиболее подвержены риску двойного учёта эффекта процессы с множественным применением моделей, смешения с бизнес-правилами, процессы с общими КПЭ, а также процессы, где симуляция работы процессов без ИИ-модели невозможна. Наиболее типичные области и кейсы:

Тип процесса	Пример
Маркетинг (склонность, удержание, акции)	Модель склонности к отклику и модель удержания одновременно влияют на одного и того же клиента
Кредитование и скоринг (PD модели)	Модель PD и скоринговые модели влияют на одни и те же кредитные решения
Ценообразование и дисконтная политика	Модели ценообразования по кредитам, депозитам, кэшбэку одновременно влияют на LTV клиента
Программы лояльности и клиентский опыт	Модель прогнозирования оттока и модель прогноза кросс-продаж влияют на одну и ту же дополнительную выручку

Примеры:

Модели продуктовых склонностей используются финансовыми организациями во всех бизнес-линиях. При этом оценка финансового эффекта на уровне одной модели сопряжена с большими затратами на изоляцию воздействия на клиента исключительно одной модели при сохранении остальных критериев неизменными. В данном случае оценка происходит на уровне выше - на уровне алгоритма оптимизации (использующего все остальные модели склонности) и предложения лучшего продуктового оффера клиенту.

Модели оценки вероятности дефолта имеют многомодульную разветвлённую структуру и обращаются при оценке кредитоспособности клиента к множеству источников информации. Оценка финансового эффекта каждого отдельного модуля модели не производится, однако производится оценка и пилотирование эффекта на уровне кредитного процесса с точки зрения влияния на выдачи, уровень риска и пр. бизнес-метрики

Процессы, наименее подверженные риску двойного учета

Существует класс процессов, в котором риски возникновения двойного учета при расчёте финансового эффекта находятся на низком уровне. Такие процессы характеризуются присутствием одиночно функционирующих моделей с низким уровнем взаимодействия с другими моделями или ограниченным влиянием на общие КПЭ. Примерами таких процессов и моделей могут служить:

- Скоринг резюме кандидатов
- Распознавание документов
- Биометрическая идентификация

Иными словами - процессы с минимальным использованием ИИ, в которых модели работают независимо.

В вышеуказанных процессах, как правило используется ограниченное число моделей, и логика их изоляции в процессе является достаточно очевидной.

Пример:

В скоринге резюме возможно выделить целевую группу, использующую результаты модели для поиска, и контрольную группу, работающую по стандартному процессу. В этом случае оценка финансового эффекта может быть получена относительно легко и без риска получения двойного учета.

ТОП-ошибок при расчете эффекта

01 - Учет только доходной части и игнорирование расходов

Пример

Внедряем персонализацию ставок кэшбэка на внутренние продукты, нацеленную на повышение утилизаций этих продуктов. В качестве эффекта считаем только прирост утилизаций, но при этом игнорируем рост затрат на выплату кэшбэка

02 - Оценка «виртуальной» экономии

Пример

С целью персонализации продукта создали 10 тыс изображений с помощью генеративной модели и добавили их в продукт. Данное изменение не принесло значимого роста в PnL продукта. Когда оценивали экономический эффект от генеративной модели оценили экономию как разницу в затратах на генерацию картинок относительно кейса, когда их бы рисовали дизайнеры. Проблема в том, что генерация изображений внутренней моделью является почти бесплатной и из-за этого может присутствовать избыточное использование инструмента. Если бы модели не существовало, то в продукте не стали бы тратить ресурс дизайнеров на создание 10 тыс. изображений.

03 - Неправильный выбор бейзлайна

Создается новый процесс в продукте с одновременным внедрением ИИ в этот процесс. Экономический эффект от ИИ оценивается относительно состояния продукт до внедрения нового процесса

Пример

Обзвон клиентов с брошенными заявками с помощью ИИ-бота.

04 - Независимая оценка взаимосвязанных моделей (игнорирование каннибализации)

2 ИИ продукта могут быть сильно скоррелированы друг с другом. Экономический эффект двух этих продуктов оценивается независимо относительно бейзлайна

с выключением конкретного продукта. В итоговом результате оба эффекта просто суммируются. В данном кейсе эффект $A+B \neq A$ и B

05 - Пренебрежение долгосрочными эффектами

Пример

Внедрение чат-бота клиентской поддержки может влиять на частотность обращений и на саму обращаемость клиентов в будущем, но экономический эффект считается только как средняя экономия времени оператора на одно случившееся обращение

06 - Ошибки в проведении A/B-тестов

Негомогенные тестовая и контрольная группы, недостаточность размера выборки, применение неправильных стат. критериев, подглядывание, отсутствие поправок на множественное сравнение

07 - Оценка неправильного эффекта

Пример

NLP-модель, которая ищет инсайты в скриптах сейлзов. Эффективность модели оценивается через тестирование применения найденных инсайтов в продажах. Однако те же самые инсайты находятся текущими сотрудниками, но дольше и дороже. Тогда нужно оценивать экономический эффект как экономию времени сотрудников.

08 - Ошибочно выбран драйвер. В рамках A/B теста выбраны не все драйверы, на которые оказывается влияние

С точки зрения банковской практики проблема возникает, если при влиянии на продажи продукта, не будет учтена продуктовая и межканальная каннибализация.

Пример не из банковской практики

Кейс:

В сервисе вызова такси внедряется новая «умная» маршрутизация, которая должна снизить среднее время ожидания машины на Y секунд. Был проведен A/B тест со стратификацией по регионам, в каждом регионе до пилотирования были проверены КГ и ЦГ на однородность, был рассчитан срок пилота, учтены все контрольные переменные – от типа машины до возраста и стажа водителя. Пилот показал реальное снижение среднего срока ожидания машины на X секунд.

Проект был растражирован. Компания стала терять огромные деньги. Почему?

Причина в том, что средний срок ожидания машины – не единственный драйвер. Новый маршрутизатор увеличил «хвосты» распределения. Иными словами, средний пользователь стал ждать машину меньше, в то время как пользователи, что ждали машину долго, стали ждать ее еще дольше, и стали уходить в другие сервисы такси. Кейс показывает, что нужно всегда смотреть на картину бизнес-процесса целиком, искать драйверы, не останавливаясь на одном, или не усредняя значение.

Отдельно хотелось бы подсветить ошибок мышления (cognitive biases), о которых необходимо помнить в процессе внедрения изменений и их оценки.

1. Эффект снятия сливок

Проблема, при которой компании выгодно работать с наиболее «удачными» с точки зрения доходности результатами/информацией/клиентами, игнорируя менее перспективные направления.

- Такая проблема может встретиться в процессе проведения пилотов – наблюдаемое изменение в первые недели может быть необоснованно большим, статистика покажет прирост операционного драйвера, кратно превышающий таргет.

В результате возникнет желание досрочно завершить пилот, что будет ошибкой, так как если посмотреть на горизонте 2-3 месяцев дельта эффекта между группами будет гораздо меньше. При проведении пилотов необходим расчет MDE, и необходимо выдерживать необходимый срок проведения.

- Эффект снятия сливок может выражаться в некорректном построении групп при пилоте (А/Б тесте), в результате которого производится ранжирование клиентов по качественному показателю, например потенциальному доходу по продукту.

Затем в Целевой Группе производится модельная доп. коммуникация именно по клиентам с высоким потенциалом, в то время как клиенты с низким потенциалом остаются в Контрольной Группе, без доп. воздействия. В результате в фин. эффект замешивается как искомый эффект от модельной коммуникации, так и эффект разницы в потенциальном доходе между группами, который существовал изначально.

2. Проблемы Морального риска (Moral hazard) и Отрицательного отбора (Adverse selection)

Это проблемы поведения клиента или сотрудника после нашего воздействия, изменения процесса, при котором мы наблюдаем обратный эффект тому, к которому мы стремились.

- Отрицательный отбор может быть, например, при определении цены услуги или продукта для клиента в зависимости от его индивидуальных особенностей.

Стремясь заложить в ставку для клиента, например, вероятность досрочного гашения кредита и компенсировать недополученный доход, на самом деле мы

можем этим простимулировать его закрыть договор еще раньше, рефинансировав кредит по меньшей ставке в другом банке, а значит банк недополучит еще больше дохода

- Еще одним примером отрицательного отбора является закладывание в цену страховки или кредита некорректной надбавки за риск (следствие ошибок первого и второго рода)

Это может привести к тому, что на такие условия, по сути нерыночные для «Хорошего» клиента, будут соглашаться именно «Плохие клиенты», которые из-за своего неблагоприятного состояния готовы на любые условия, а их риск в итоге реализуется с высокой вероятностью

- Примером же морального риска может быть поведение сотрудника продаж после перехода от контроля операций руководителем к проверкам с помощью ИИ алгоритмам

Сотрудник может расслабиться, в надежде, что ИИ алгоритм отследит и не позволит совершить ему любые ошибки, в отличие от руководителя-человек. В результате есть риск, наоборот, значительного увеличения количество ошибок сотрудника, так как ИИ алгоритм обучался при условиях и подразумевает базовую осторожность сотрудника.

3. Confirmation Bias (Предвзятость подтверждения) - стремление человека к доказательству собственного мнения.

Это стремление состоит из трех аспектов:

- Предвзятый поиск информации: люди при поиске информации для модели/инициативы стремятся игнорировать «неудобную» информацию, которая может исказить оценку в неблагоприятную для эффекта сторону.
- Предвзятая интерпретация: люди, которым нужно проинтерпретировать собранную информацию, будут стремиться субъективно «подогнать» описание под желаемую теорию, игнорируя «неблагоприятные» интерпретации.
- Предвзятость памяти: люди, стремящиеся доказать свою точку зрения, будут вспоминать и использовать доказательства в свою пользу.

Такое поведение, чрезмерное стремление любыми способами доказать свою правоту, приводит к смещению результатов исследований, нарушению принципа объективности и принятию ошибочных для бизнеса компании решений.

Кейсы с примерами оценки фин. эффектов от отдельных внедрений



182 млн. руб.
за 2023 г.

КЕЙС 1:

Кампании продаж и модели приоритизации клиентов на основе ИИ

Основные факты

- Ключевые бизнес-метрики
Продуктовая конверсия
- Тип эффекта
Эффект на доход
- Сегмент клиентов:
B2B
- Сегмент клиентов:
B2B
- Метод оценки фин эффекта
Рандомизированный А/Б тест
- Тип изменения
Автоматизация
- Описание Технологии
Классификационные модели (CatBoost, и т.п.)
- Описание Технологии
Классификационные модели (CatBoost, и т.п.)

Описание

Цель модели

Ранжировать продуктовые предложения по итогу сессии взаимодействия клиента с продуктами на сайте sberbank.ru / в личном кабинете Сбербизнес.

Бизнес цель

Увеличить процент совершения сделок по продуктам, с которыми наиболее глубоко взаимодействовал клиент, перевести “брошенные” заявки в заключенные сделки.

Пример расчета

Калькулятор:

$$\text{Финансовый Эффект}_T = \sum_{j=1}^T \sum_{i=1}^N \text{ДельтаКонверсии}_i * \text{БазаЛидов}_j * \text{PL}_{ij}$$

где Т – месяц, к которому проводится оценка от инициативы; N – количество продуктов сплита; ДельтаКонверсии_i – это дельта конверсии i-ого продукта сплита, полученная по пилоту; БазаЛидов_j – это база лидов в месяц j; PL_{ij} – это значения PL i-ого продукта в j-ый месяц жизни этого продукта. При этом, поскольку пилот проводился только 3 месяца (нет постоянного контроля), на значение ДельтаКонверсии_i выставляется мониторинг, корректирующий значения конверсии в случае необходимости.

КЕЙС 2: Чат-бот клиентского обслуживания

Основные факты

- Ключевые бизнес-метрики
Время операторов поддержки
- Тип эффекта
Расходы
- Сегмент клиентов:
B2C
- Периметр внедрения кейса
Вся активная клиентская база Т-банка
- Метод оценки фин эффекта
А/Б-тест с вечным контролем (holdout)
- Тип изменения
Автоматизация
- Описание Технологии
NLP
- Описание бизнес задачи
Чат-бот клиентского обслуживания:
распознает намерения клиента с помощью ML и дает релевантные ответы без участия человека, может предоставить информацию или решить запрос прямо в чате.

Описание

Расчет экономического эффекта осуществляется с помощью АБ-теста с вечным контролем относительно альтернативного сценария работы поддержки без чат-бота.

Дизайн эксперимента:

- Сплит по клиентам при обращении в чат поддержки.
- Контроль (1%) - боты отключены, клиент попадает сразу на операторов;
- Тест (99%) - сначала помогает бот

Пример расчета

Калькулятор:

$$\text{delta } t = \left(\frac{\text{время операторов за месяц в контроле}}{\text{кол-во клиентов в контроле}} - \frac{\text{время операторов за месяц в тесте}}{\text{кол-во клиентов в тесте}} \right) * \text{кол-во клиентов в тесте} =$$

= время операторов за месяц в контроле * 99 - время операторов за месяц в тесте = 2,11 млрд. сек.

$$\text{Экономия} = \text{delta } t * \text{стоимость секунды оператора} = 2,11 \text{ млрд. сек.} * 0.205 \text{ руб/с} = 433 \text{ млн. руб. в мес.}$$

КЕЙС 3:

Модель определения дохода клиента в целях расчета Показателя Долговой Нагрузки

Получили премию GIA 2024 – Ежегодная премия в области корпоративных инноваций

<https://gia.generation-startup.ru/>

Основные факты

- Ключевые бизнес-метрики eROE, потребление капитала
- Тип эффекта Эффект на доход
- Сегмент клиентов: B2C
- Метод оценки фин эффекта A/B тестирование
- Тип изменения Автоматизация
- Описание Технологии ML (градиентный бустинг – утверждено с ЦБ)
- Периметр внедрения кейса Десятки миллионов клиентов (клиентская база АЛЬФА-БАНКА)
- Описание бизнес задачи Расчет Показателя Долговой Нагрузки при помощи модельного подхода к оценке дохода клиента

Описание

В 2023 г. АО «АЛЬФА-БАНК» в тройке первых получил разрешение от Банка России о допустимости применения собственных моделей для расчета Показателя долговой нагрузки, в 2025 г. получил разрешение на обновленную модель. Благодаря разработанным моделям АО «АЛЬФА-БАНК» более персонализировано предлагает свои продукты, более точно просчитывает риски невозврата, при этом улучшает клиентский путь.

Пример расчета

Калькулятор:

$$eROE = \frac{\text{ожидаемая чистая прибыль}}{\text{потребление капитала}} = \frac{\text{ожидаемая чистая прибыль}}{RWA * CAR}$$

$$eROE_{\text{без модели}} = \frac{10\%}{2.5 * 11.5\%} = 35\%$$

$$eROE_{\text{с моделью}} = \frac{10\%}{1.5 * 11.5\%} = 58\%$$

$$\Delta eROE = eROE_{\text{с моделью}} - eROE_{\text{без модели}} = 23\%$$

eROE – ожидаемая доходность

RWA – активы, взвешенные по уровню риска

CAR – норматив достаточности капитала

Чистая прибыль ↑, объемы бизнеса ↑, удовлетворенность клиентов ↑, потребление капитала ↓

КЕЙС 4:

Прогнозирование метрик регулярной отчетности
с помощью AutoML

Основные факты

- Ключевые бизнес-метрики
MAPE, трудозатраты
- Тип эффекта
Потенциальные расходы
- Сегмент клиентов:
Внутренний клиент
- Метод оценки фин эффекта
Моделирование
- Тип изменения
Автоматизация
- Описание Технологии
AutoML

Описание

Для повышения качества прогнозирования в регулярной отчетности, необходимо замещение классических статистических методов прогнозирования (в т.ч. run-rate) на прогнозирование методами машинного обучения, что улучшает точность прогноза по каждой метрике от 5% до 60 раз. Т.к. прогнозировать нужно более 1500 метрик, то при еженедельной периодичности прогноза, при регулярном re-feet моделей, требуется дополнительное привлечение огромного штата дата-саентистов. Использование AutoML решает эту проблему. При подаче данных происходит формирование объясняющих факторов, отбор метода ML из 378 (учитывая стекинг) и автоматический фиче-инжиниринг (отбор факторов). На выходе бизнес-аналитик автоматически получает модель ML и прогноз. По сути инструмент автоматизирует работу дата-саентистов, которые являются достаточно дорогими сотрудниками.

Пример расчета

Калькулятор:

Фин.эффект = 1500 моделей X 5 чел.-дней X 52 расчётов/год = 390 тыс.чел.-дней./год = 1611 ПШЕ / год = 10,67 млрд.руб./год

Трудозатраты на реализацию = 80 чел.-дней = 2,2 млн.руб. ; Трудозатраты на поддержку = 40 чел.-дней/год = 1,1 млн.руб./год.;

$ROI^* = (10,67 \text{ млрд.руб./год} \cdot 3 \text{ года} - 1,1 \text{ млн.руб./год} \cdot 3 \text{ года} - 2,2 \text{ млн.руб.}) / (1,1 \text{ млн.руб./год} \cdot 3 \text{ года} + 2,2 \text{ млн.руб.}) = 581900\%$

* при расчёте на 3 года

КЕЙС 5:

AI-агент по формированию one-page-слайда по презентации

Основные факты

- Ключевые бизнес-метрики
Трудозатраты
- Тип эффекта
Потенциальные расходы
- Сегмент клиентов:
Внутренний клиент
- Метод оценки фин эффекта
Моделирование
- Тип изменения
Автоматизация
- Описание Технологии
LLM, AI-агент
- Периметр внедрения кейса
Блок финансы
- Описание бизнес задачи
Формирование one-page

Описание

Бизнес аналитикам приходится периодически формировать one-page слайд на основе большой презентации. Это долго и трудозатратно. Использование AI-агента для формирования one-page сокращает трудозатраты с многих часов до нескольких секунд. Агент принимает на вход презентацию pptx и преобразует элементы презентации (текст, графики, таблицы, диаграммы, картинки) в формат удобный для LLM. Далее LLM делает саммари материала и выдает результат в определенном xml-подобном формате. Далее агент через специализированные python библиотеки формирует из этого «xml» слайд презентации pptx.

Пример расчета

Калькулятор:

Фин.эффект = 1 рабочий день X 50 расчётов/год = 50 дней./год = 0,2 ПШЕ / год

Трудозатраты на реализацию = 40 рабочих дней; Трудозатраты на поддержку = 20 рабочих дней;

$ROI^* = (50 \cdot 3 - 20 \cdot 3 - 40) / (20 \cdot 3 + 40) = 50\%$

КЕЙС 6:

Скоринг для исходящих коммуникаций с клиентами банка в Контактном центре

Основные факты

- Ключевые бизнес-метрики
CR/NPV/PnL
- Тип эффекта
Доходы/оптимизация расходов
- Сегмент клиентов:
B2C (клиенты банка)
- Метод оценки фин эффекта
A/B тестирование: ML против случайной выборки
- Тип изменения
Оптимизация
- Описание Технологии
ML (модели классификации)
- Периметр внедрения кейса
Рекомендации по продуктам Банка в КЦ
- Описание бизнес задачи
Персональные рекомендации, оптимизация конверсии по продажам, повышение эффективности КЦ

Описание

Разработаны ML-модели оценки вероятности отклика на продукты Банка после коммуникации. Клиенты с высокой вероятностью отклика на продукт попадали в списки для исходящих звонков.

Пример расчета

Калькулятор:

$$\text{ФинЭффект} = \sum (\text{СегментКлиентов}_{ML} \times (CR_{ML} - CR_{control}) * PNL_{ii} - E_i) - E_{ML}$$

CR_{ML} - отклик в группе отобранной по модели, $CR_{control}$ - отклик в случайно отобранной группе

PNL_{ii} - PnL за один год по продукту i

E_i - затраты на проведение коммуникаций по продукту i , включая ЦГ и КГ

E_{ML} - затраты на инфраструктуру и ФОТ для разработки ML модели

КЕЙС 7: Создание AI-аватара для сотрудников поддержки

Основные факты

- Ключевые бизнес-метрики
CSAT, eNPS
- Тип эффекта
Расходы
- Сегмент клиентов:
B2B / B2C / B2E
- Метод оценки фин эффекта
A/B тестирование, месяц
- Тип изменения
Реальные фото → Генеративные
- Описание Технологии
GenAI (Диффузии) — Remini и Magic Eraser
- Периметр внедрения кейса
500 операторов
- Описание бизнес задачи
Создание однотипных аватарок для специалистов поддержки с помощью GenAI-сервисов

Описание

Мы разработали решение — ИИ-аватарки для сотрудников поддержки. Ранее каждый выбирал фото самостоятельно, что снижало единый стиль и доверие. Новая технология создаёт уникальные аватарки, решая проблему отсутствия фото и улучшая коммуникацию с клиентами. Кроме того, аватарки защищают сотрудников, исключая возможность поиска их в соцсетях недобросовестными клиентами. Внедрение повысило вовлечённость сотрудников и улучшило показатели удовлетворённости клиентов (CSAT). Проект объединяет инновации и заботу о клиентах и сотрудниках, повышая качество обслуживания и лояльность.

Пример расчета

Калькулятор расчета эффекта **среднее за месяц:**

Рост положительных оценок =

$$\frac{(\text{Новые положительные оценки за месяц после} - \text{Исходные положительные оценки за месяц до})}{\text{Исходные хорошие оценки}} \times 100\%$$

Исходные оценки: 6,400

Новые оценки: 6,976

Прирост = $(6,976 - 6,400) / 6,400 \times 100\% = 9\%$

КЕЙС 8:

Чат и голосовой бот клиентского обслуживания

Основные факты

- Ключевые бизнес-метрики
PnL
- Тип эффекта
Расходы
- Сегмент клиентов:
B2C
- Метод оценки фин эффекта
Серия A/B-тестов без вечного контроля
- Тип изменения
Автоматизация
- Описание Технологии
NLP
- Периметр внедрения кейса
Активная клиентская база физических лиц банка
- Описание бизнес задачи
Цель — сократить совокупное время операторов, сохранив или улучшив удовлетворённость клиентов, тем самым снизив прямые операционные расходы.

Описание

Чат- и голосовой бот:

- распознаёт намерение клиента (NLP),
- с помощью большой языковой модели выбирает наилучший ответ в рамках темы
- полностью закрывает запрос без участия человека.

Частично автоматизированные диалоги (когда оператор подключается) не включаются в дальнейший расчёт эффекта.

Пример расчета

Калькулятор расчета эффекта за год:

$$1. \text{ Доля полностью автоматизированных диалогов: } p = \frac{N_{\text{auto}}}{N_{\text{total}}}$$

$$2. \text{ Сэкономленное время операторов: } \Delta T = p \times N_{\text{total}} \times \text{top}$$

$$3. \text{ Экономия в деньгах: Экономия} = \Delta T \times C_{\text{сек}}$$

N_{auto} – диалоги, полностью закрыты ботом

N_{total} – все диалоги за период

top – среднее время оператора на 1 диалог, сек

$C_{\text{сек}}$ – себестоимость секунды работы оператора, руб

Экономия
~ 410 млн руб. год



КЕЙС 9: Модель привлечения BARCELONA (продукт Премиум)

Основные факты

- Ключевые бизнес-метрики
CR, CTR
- Тип эффекта
Эффект на доход
- Сегмент клиентов:
B2C
- Метод оценки фин эффекта
A/B тестирование
- Тип изменения
Автоматизация
- Описание Технологии
GBM модель
- Периметр внедрения кейса
CRM система розничного бизнеса
- Описание бизнес задачи
Привлечение новых клиентов
на пакет услуг Премиум

Описание

Модель склонности к оформлению пакета услуг Премиум на основе истории откликов клиентов

Пример расчета

Калькулятор:

Кол-во привлеченных клиентов * LTV клиента = Итоговый эффект

КЕЙС 10:

Аналитический сервис по автоматизации процесса оценки и переоценки отдельных видов профильных залогов юридических лиц (БОАИ - база отраслевой экономической информации)

Основные факты

- Ключевые бизнес-метрики
коэффициент покрытия
объектов типового залога
- Тип эффекта
Доходы
- Сегмент клиентов:
B2B
- Метод оценки фин эффекта
A/B-тестирование
- Тип изменения
Автоматизация
- Описание Технологии
ML (градиентный бустинг и линейные модели)
- Периметр внедрения кейса
Внутреннее подразделение Банка, ЮЛ
- Описание бизнес задачи
Оценка залоговой стоимости имущества ЮЛ

Описание

Автоматизация процесса оценки и переоценки отдельных видов профильных залогов клиентов-юридических лиц за счет внедрения аналитического сервиса, использующего технологии ИИ

Цель реализации

1. Ускорение процесса определения стоимости отдельных видов профильного залогового обеспечения до онлайн оценки
2. Сокращение прямых затрат Банка и повышение объективности оценки отдельных видов профильного залогового обеспечения
3. Обеспечение возможности мониторинга стоимости активов на портфельном уровне

Пример расчета

Калькулятор:

Применимость (покрытие) сервиса по оценке залогов – 50% номенклатуры.

Вычисляется как отношение числа объектов залога, по которому сервис обеспечил переоценку в границах доверительного диапазона (+/-10% от предыдущей оценки, проведенной экспертом Банка), к общему числу объектов залога тех видов, к которым применим сервис (транспорта, техника, прицепное и навесное оборудование)

Ожидаемый доход к 2029 г./млн руб. = ежеквартальное сокращение расходов на содержание персонала (заработная плата и страховые взносы (ОМС, ОПС, ОСС) + выплаты при увольнении) X 4 квартала X количество лет реализации бизнес-кейса (19,6928+0)X4X6=472,62

КЕЙС 11:

Время на поиск ответов по бизнес Википедии и продуктовому каталогу для сотрудников колл центра (КЦ)

Основные факты

- Ключевые бизнес-метрики
Average Usefulness (AU)
- Тип эффекта
Эффект на расходы
- Сегмент клиентов:
Retail
- Метод оценки фин эффекта
Измерения
- Тип изменения
Автоматизация
- Описание Технологии
LLM (RAG сервис)
- Периметр внедрения кейса
Сотрудники КЦ (> 100 чел)
- Описание бизнес задачи
Персональные рекомендации

Описание

Оптимизация трудозатрат от сокращения времени на поиск ответов сотрудниками КЦ

Пример расчета

Калькулятор:

$$E = \sum_i (N(i) * (AT_Control - T1_Pilot) * CostPer1min(i))$$

$N(i)$ Количество звонков в КЦ, обрабатываемых операторами
(в пилотной группе, по годам i)

$AT_Control$ Среднее время телефонного диалога до реализации проекта
(на основании данных журналирования)

$T1_Pilot$ Среднее время телефонного диалога после реализации проекта
(на основании расчетов SLA сервиса LLM+RAG)

$CostPer1min(i)$ Стоимость 1 минуты диалога Оператора КЦ
(с учетом прогноза роста по годам, i)

