

Как не дать чат-боту разочаровать клиента:

оперативный анализ и управление качеством ответов
диалоговых ассистентов на основе генеративного ИИ

Так ли нужны бизнесу чат-боты с ИИ?

68%

экспертов по внутренним процессам считают, что ИИ и чат-боты приведут к большей экономии средств в ближайшие годы.

[Zendesk. CX Trends Reports](#)

80%

руководителей банков считают, что повышение ценности бизнеса с помощью ИИ поможет отличить победителей от проигравших.

[Опрос The Economist](#)

61%

Представителей бизнеса считают, что чат-боты на основе ИИ могут значительно повысить производительность за счет автоматизации типичных задач.

[Artsmart](#)

70%

коэффициент конверсии чат-ботов на основе ИИ в некоторых секторах

67%

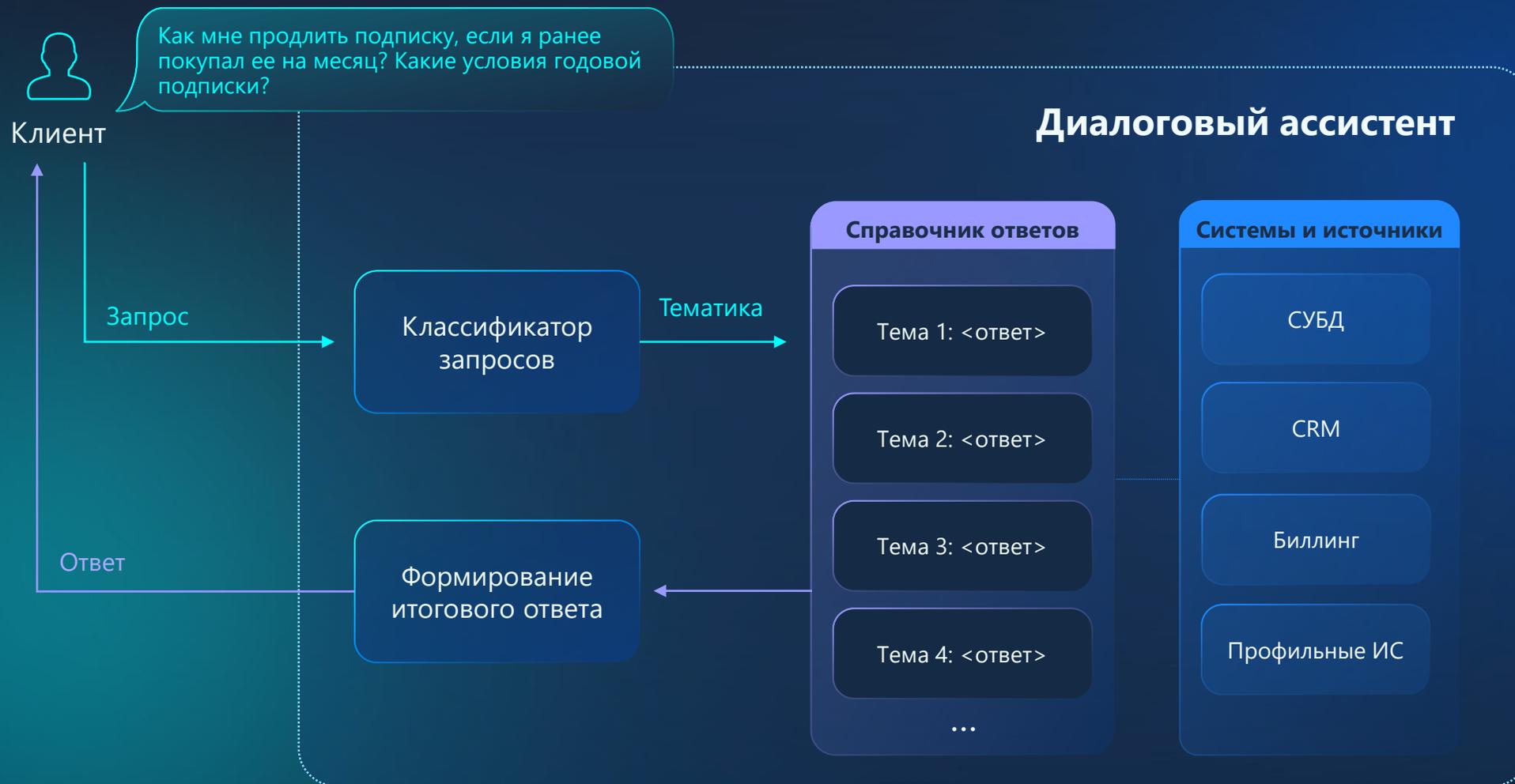
Средний рост продаж благодаря использованию чат-ботов на базе ИИ по мнению руководителей предприятий

[Artsmart](#)

Главные вопросы о качестве работы ботов с ИИ

- 01 Как изменилась аналитика диалогов виртуального ассистента после перехода на модель с GenAI*?
- 02 Возможно ли контролировать точность генеративного ответа прямо во время диалога?
- 03 Как использовать и внедрять автооценку модели в диалоговых ассистентах?
- 04 Как аналитику интерпретировать результаты автооценки для управления качеством ответов?
- 05 Как автоматизация и оперативный контроль помогают диалоговому ассистенту на базе GenAI стать полезным собеседником?

Модель диалогового робота pre-GenAI



Аналитика диалогов pre-GenAI

Что оцениваем?

- Точность классификации
- Полнота ответов
- Корректность логики диалогового ассистента

Действия по итогам оценки

- Обучить классификатор на новых данных
- Наполнить базу знаний: добавить новые ответы, внести изменения в существующие ответы
- Внести изменения в логику диалогового ассистента

Вся информация под контролем, но ошибка классификации вносит риск, а диалог не подстроен под реальный запрос клиента *(ответ формально корректный, но не отвечает на вопрос клиента)*



до 100

диалогов в сутки в среднем оценивает один сотрудник*

Модель диалогового робота на базе GenAI&RAG



Как оценивать генеративные ответы

Новые риски

- ➔ **Нет фиксированных**, заранее заданных ответов – потребуется дополнительное время аналитика на проверку ответов
- ➔ **Нет эталонных ответов** – с чем сравнить фактический ответ робота?
- ➔ **По каким критериям** оценивать ответ робота?
- ➔ **Где искать ошибку?** Точки риска: этап поиска, этап формирования контекста, этап генерации
- ➔ **Главная опасность** – предоставить клиенту ложную информацию. Как оперативно выявить такие случаи?

Кажется, что контроля становится меньше...

Что найдено в базе знаний?

Что сгенерировал GenAI?

До 100

диалогов в сутки в среднем оценивал один сотрудник в pre-GenAI схеме

Сколько диалогов с генеративными ответами оценит сотрудник, не имея вспомогательных инструментов?

Оперативный контроль: онлайн и офлайн сценарий

Онлайн-режим (во время диалога)

Генерация
ответа

Самооценка
GenAI

Принятие
решения об ответе

Сохранение
данных метрик

Вопрос: Есть ли возможность отменить подписку не насовсем, а на время?

Ответ ИИ: Да, такая возможность существует. Это называется "заморозкой" подписки. Вы можете временно приостановить действие своей мультиподписки, сохранив при этом все неистраченные дни, а затем возобновить ее в любой удобный для вас момент. Как это сделать: - Войдите в свой личный кабинет. - Выберите необходимый тариф. - Нажмите кнопку «Заморозить». Обычно заморозка доступна на 7 дней, однако в рамках персональных предложений срок может быть увеличен до 31 декабря 2025 года.

Релевантность: 85%
Достоверность: 98%
Полнота: 36%
Финальная оценка: 73%

Внесение
изменений

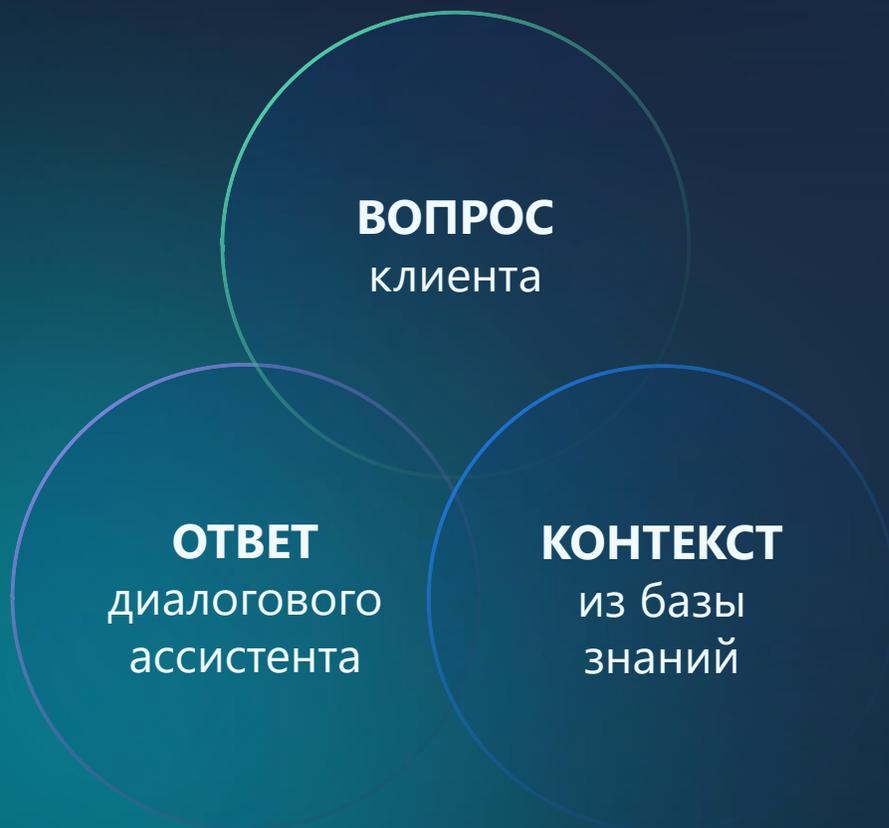
Выявление
ошибок

Анализ в
разрезе метрик

Формирование
статистики

Офлайн-режим (этап аналитики)

Триада метрик автооценки GenAI



Вопрос клиента

Генеративный ответ

Релевантность ответа:
определить, достаточно ли ОТВЕТ соответствует ВОПРОСУ

Генеративный ответ

Контекст

Полнота ответа: определить, достаточно ли фактов из КОНТЕКСТА содержится в ОТВЕТЕ

Генеративный ответ

Контекст

Достоверность ответа:
определить, соответствует ли ОТВЕТ фактам из КОНТЕКСТА

Триада метрик автооценки GenAI



$$\alpha * \text{answer_relevancy} + \beta * \text{faithfulness} + \gamma * \text{completeness}$$

Итоговая оценка — взвешенная сумма трех метрик

Сравнение итоговой оценки с пороговым значением



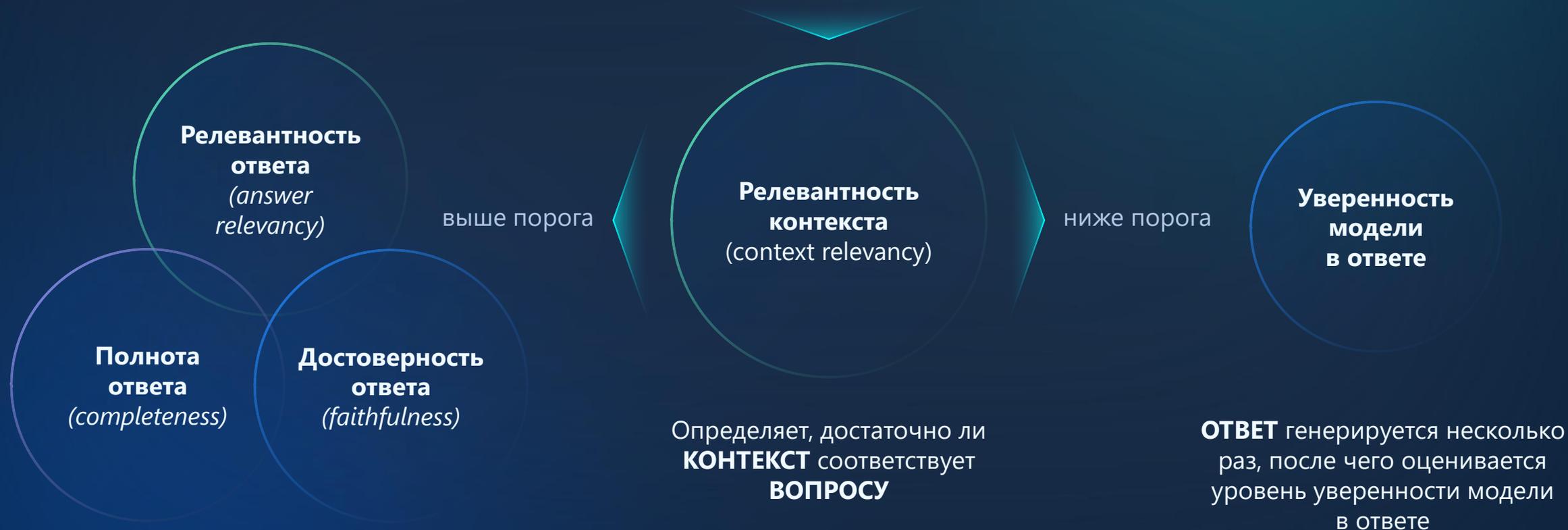
Принятие решения об отправке генеративного ответа клиенту



Формирование статистики по метрикам и оценкам

Автооценка GenAI для неклассифицированных вопросов

Контекст + вопрос клиента



Принципы алгоритма самооценки GenAI

- Генеративный ответ диалогового ассистента оценивается в режиме реального времени до отправки ответа клиенту
- Самооценка может использоваться в сценариях как с одной LLM, так с несколькими LLM для выбора наилучшего ответа
- Самооценка работает для целевых (классифицированных) и нецелевых запросов
- Все оценки представлены в одной шкале измерения (%)

Точность выбранного алгоритма:
83% для целевых вопросов и
88% для нецелевых вопросов

- Выбор метрик и подбор весов для итоговой формулы проводились по результатам сравнения с ручной экспертной оценкой ответов LLM.
- В алгоритме используется порог итоговой оценки, ниже которого генеративный ответ считается непригодным для отправки пользователю – такие случаи фиксируются как «Не выиграла ни одна модель»

Аналитика самооценок GenAI

Какая статистика нужна аналитику

- Сколько ответов прошли по порогу и были отправлены клиенту
- Сколько ответов не прошли по порогу и были непригодными из-за низкой самооценки
- При наличии в сценарии классификатора – средние/медианные итоговые оценки в разрезе каждой тематики
- Средние/медианные показатели по каждой метрике, в том числе в разрезе тематик



Аналитика самооценок GenAI

Интерпретация и анализ

«Проблемная» группа

Ниже порога: ответы неточные / неполные / нерелевантные

Оценка средних метрик
и выявление низкого показателя:
релевантность / полнота / точность

Приоритизация

Определение стратегии корректировки

Особое внимание к случаям расхождения метрик

Оценка метрик в разрезе тематик

Выявление слабых мест базы знаний

Изменения

Контроль динамики метрик

Аналитика самооценок GenAI

Взаимосвязи метрик

Сценарий	Интерпретация	Возможные причины	Действия
Высокая полнота и низкая достоверность	Робот взял много фактов, но исказил информацию	Некорректное перефразирование, «галлюцинации», противоречия в предоставленном контексте	<ul style="list-style-type: none">• Промптинг или дообучение модели• Контроль стиля генерации• Проверка базы знаний на противоречия и недостающую информацию• Ужесточение требований к цитированию контекста
Высокая достоверность и низкая релевантность	Робот сообщает правдивую информацию, но не по теме	Ошибка интерпретации вопроса	<ul style="list-style-type: none">• Улучшить поиск и отбор контекста• При генерации ответа инструктировать модель на проверку контекста
Высокая релевантность и низкая полнота	Робот понял вопрос, но не раскрыл ответ в полной мере	Недостаточный контекст, плохая агрегация фактов	<ul style="list-style-type: none">• Улучшить ранжирование результатов поиска• Поработать с объединением / суммаризацией контекста

От метрик автооценки к UX-метрикам

Релевантность ответа

- Клиент чувствует, что его поняли
- Нет раздражения от ответов «невпопад»

↑ **CSAT** – чем выше точность ответа, тем выше удовлетворенность

↑ **FCR** – уменьшается необходимость возвращаться за уточнением

↓ **CES** – снижает усилия для получения информации

Достоверность ответа

- Клиент может положиться на ответ – больше доверия к бренду или продукту

↑ **NPS** – точность и достоверность ответов повышает лояльность и доверие

↓ **Error Rate, Escalation Rate** – чем точнее ответы, тем меньше жалоб и эскалаций

Полнота ответа

- Клиент сразу получает детали
- Экономия времени

↓ **AHT** – чем меньше дополнительных вопросов, тем быстрее результат обращения

↑ **FCR** – полнота ответа повышает шансы решить вопрос с первой попытки

Спасибо за внимание!

Москва

105120
2-ой Сыромятнический пер., 1
+7 495 669 7440
stc-msk@speechpro.com

Санкт-Петербург

194044
Выборгская наб., 45, лит. Е
+7 812 325 8848
stc-spb@speechpro.com