iQSTORE
POWERED BY iCAM

ИИ вам не поможет

Или как построить бота, который реально работает

Дмитрий Лосев
Продакт-менеджер, iQStore

О компании

iQStore (ООО «АЙ КЬЮ СТОР») – российская ИТ-компания, входящая в состав крупного консорциума iCAM Group и занимающаяся разработкой программных решений для создания бесшовного клиентского сервиса. Экспертиза компании – омниканальные коммуникации и автоматизация обслуживания с применением искусственного интеллекта.

14 лет
исследований и разработок

50+
клиентов крупного бизнеса

15+
решений для бизнеса

150+
успешных проектов

Боты первой линии сменили поколение, но проблемы остались

Сценарные

кнопки и ветки сценария

Как сменить ПИН-код карты?

Выберите раздел

Карты

Вклады

NLU

распознают намерение, отвечают шаблоном

Как сменить ПИН-код карты?

ПИН-код можно сменить в банкомате или интернет-банке

А если карта цифровая?

ПИН-код можно сменить в банкомате или интернет-банке

ИИ-боты

свободный текст по базе знаний

Как сменить ПИН-код карты?

Вы можете бесплатно сменить ПИН-код в интернет-банке, в банкомате или на кассе банка. В интернет-банке новый код придёт в СМС, а в банкомате вы сможете задать его самостоятельно.

Ожидание рынка: взять модель помощнее – и бот заговорит
Реальность: без правильной архитектуры ошибается даже мощная модель

Шаг 1: простой вариант, который действительно работает

Клиент: «Как узнать CVC-код?»

База знаний

Языковая модель

Бот: «CVC-код – на обратной стороне карты»

Проверили на реальных данных:

98 вопросов клиентов банка

2 модели: малая (8B) и средняя (27B)*

107 диалогов по три хода

5 сборок бота: от 2 до 6 блоков

* 8B и 27B – 8 и 27 млрд. параметров

Разница в точности – минимальна, разница в цене – втрое

Стоимость железа

малая ×1

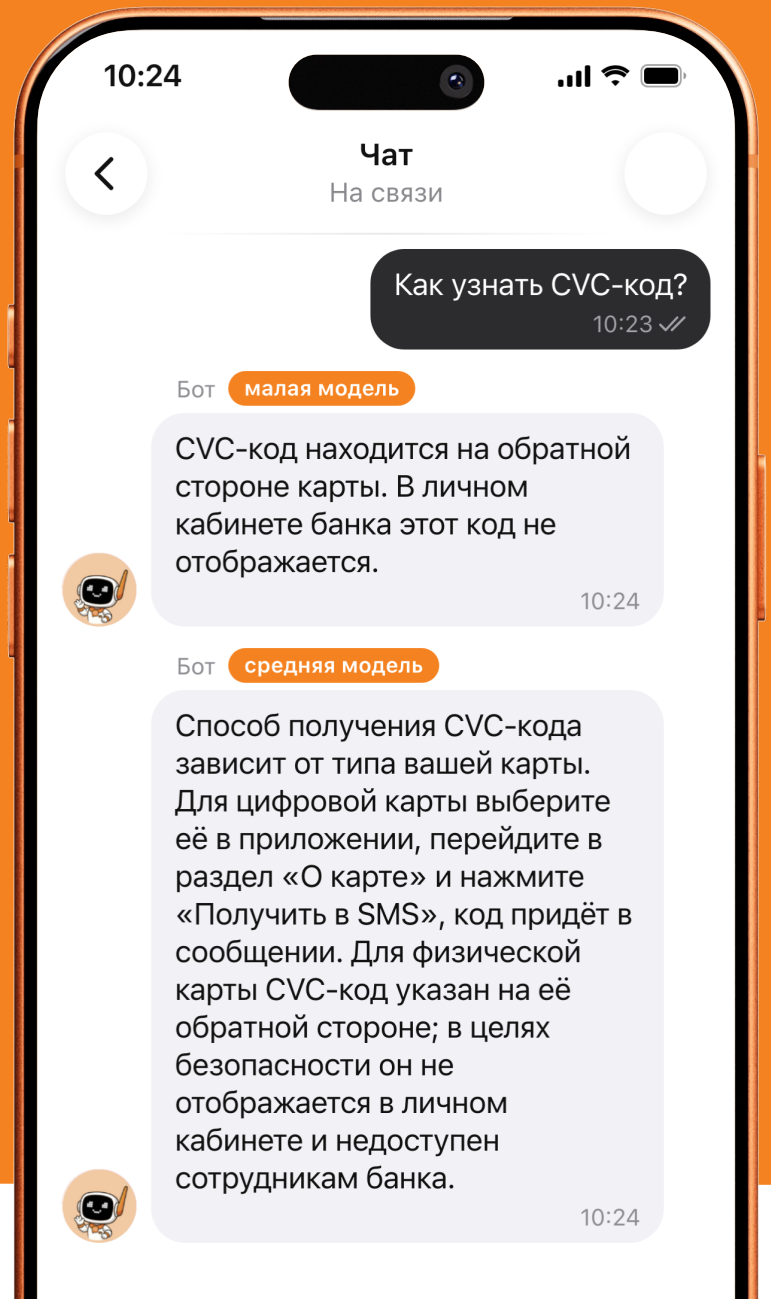
средняя ×3

Верных ответов*

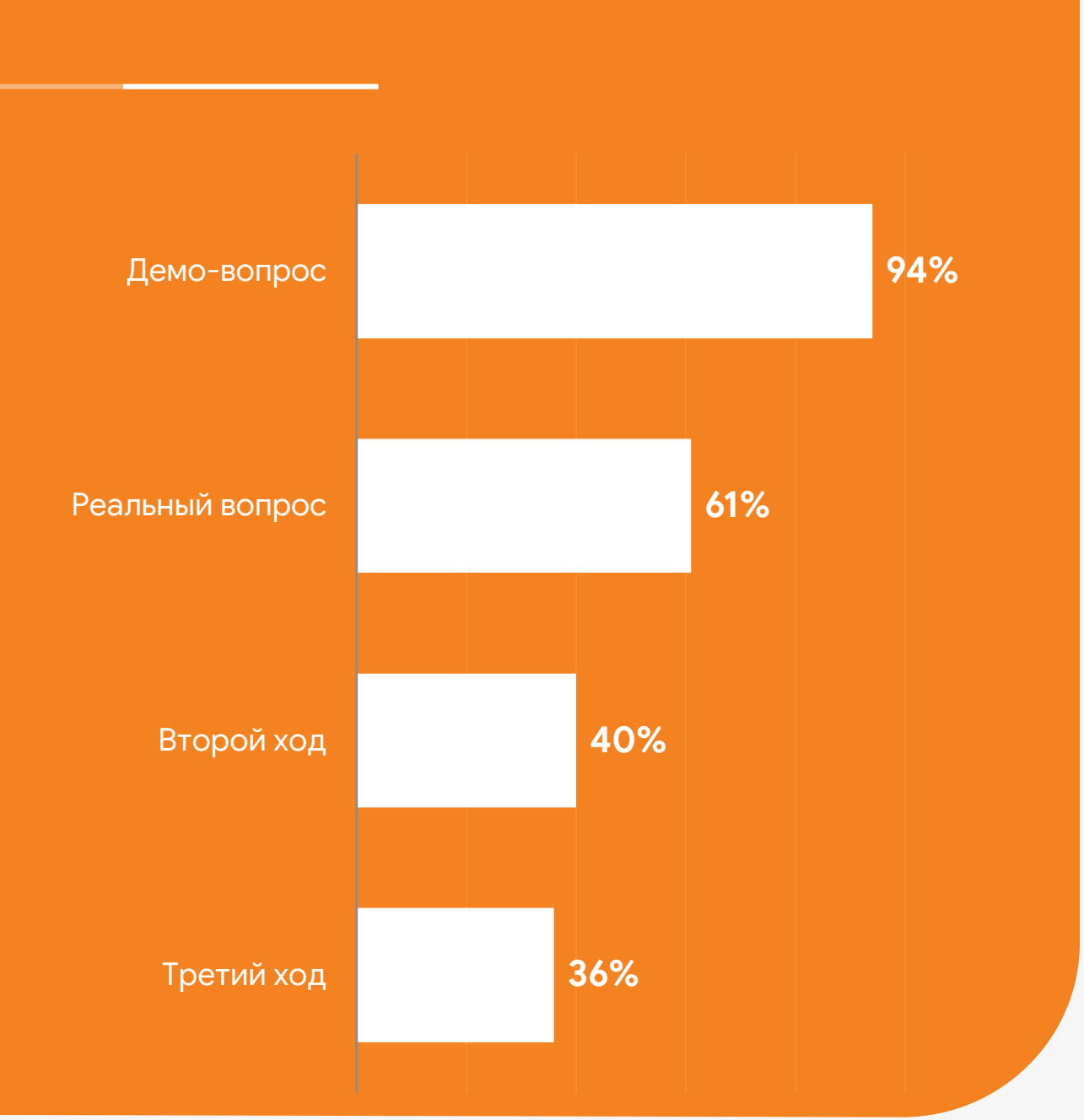
малая 55%

средняя 61%

Готовы ли вы платить за разницу, которую клиент не заметит?



* 98 реальных вопросов клиентов банка, схема «база + модель»



Почему на демо всё работает, а в реальности – нет?

Потому что клиенты:

- **Пишут своими словами**
- **Уточняют**
- **Не только спрашивают**
- **Задают опасные вопросы**

Демо-вопрос ≠ реальный клиент.

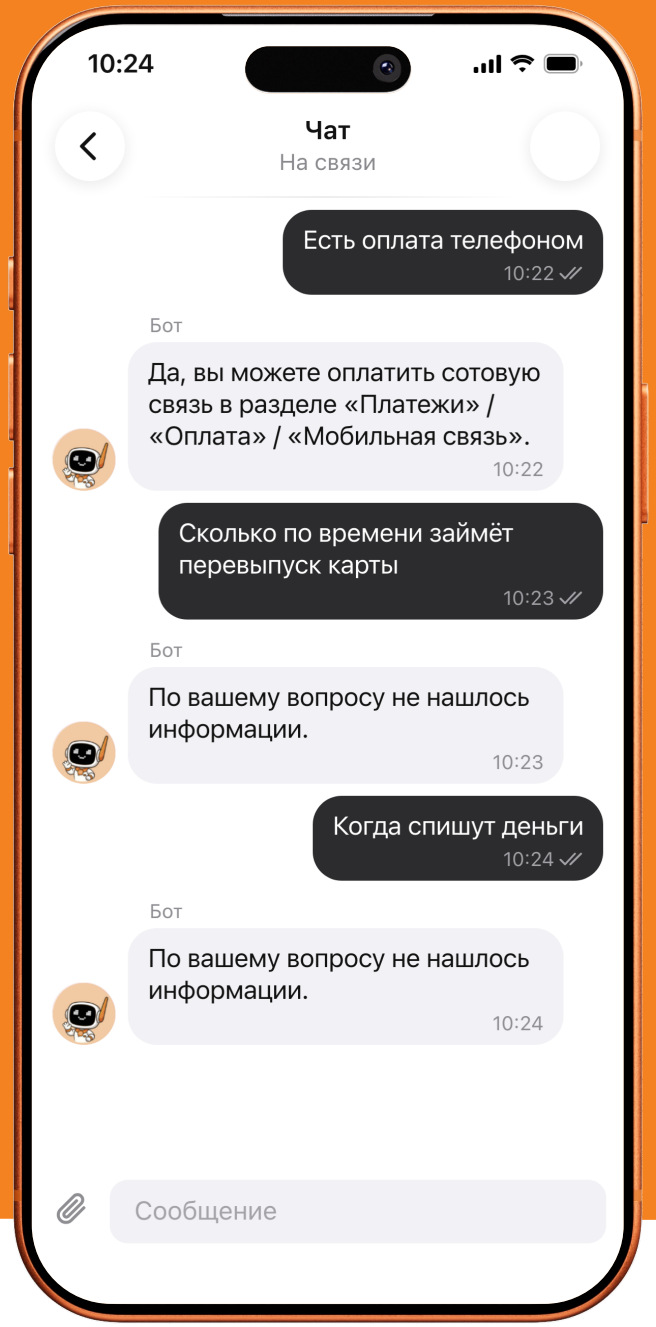
Ответ в базе есть, но бот его не находит

2 раза из 8
нужный раздел попал в выдачу

3 раза из 8
ответил «не нашлось информации»

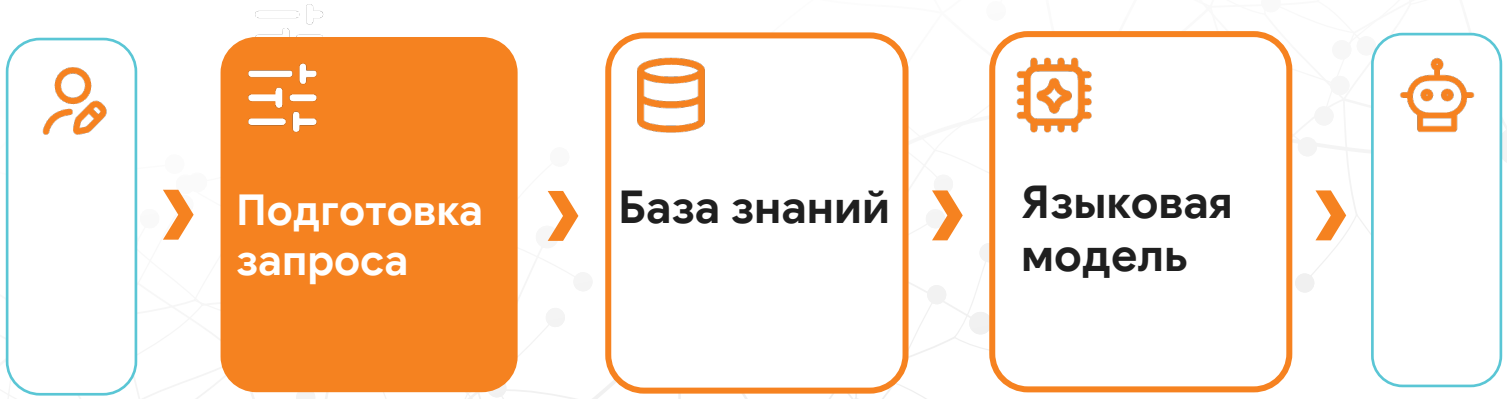
2 раза из 8
неверный ответ по другому разделу

Дело не в размере модели, а в формулировках.



Шаг 2: подготовка запроса дешевле, чем новая модель

Мы не меняем модель. Мы меняем то, что ей приносим.



Есть оплата телефоном?

Словарь терминов

+ NFC, бесконтактная оплата

Переформулировки

Настройка бесконтактной оплаты NFC
Активация оплаты смартфоном

Найден раздел
Mir Pay ✓

Подготовка запроса решила проблему, с которой не справилась средняя модель.

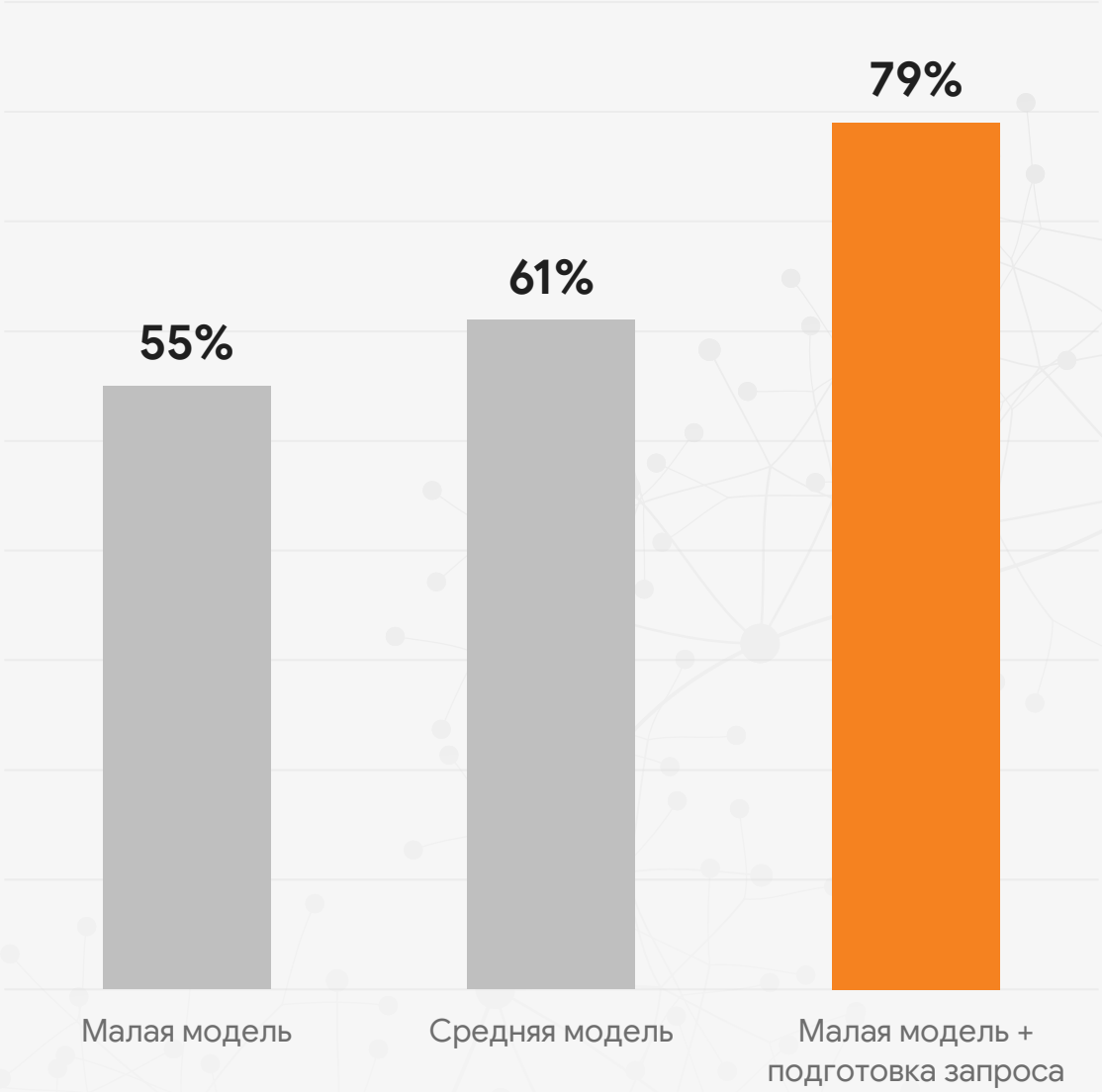
Правильные фрагменты важнее большой модели

+6 п.п.
дала модель втрое дороже

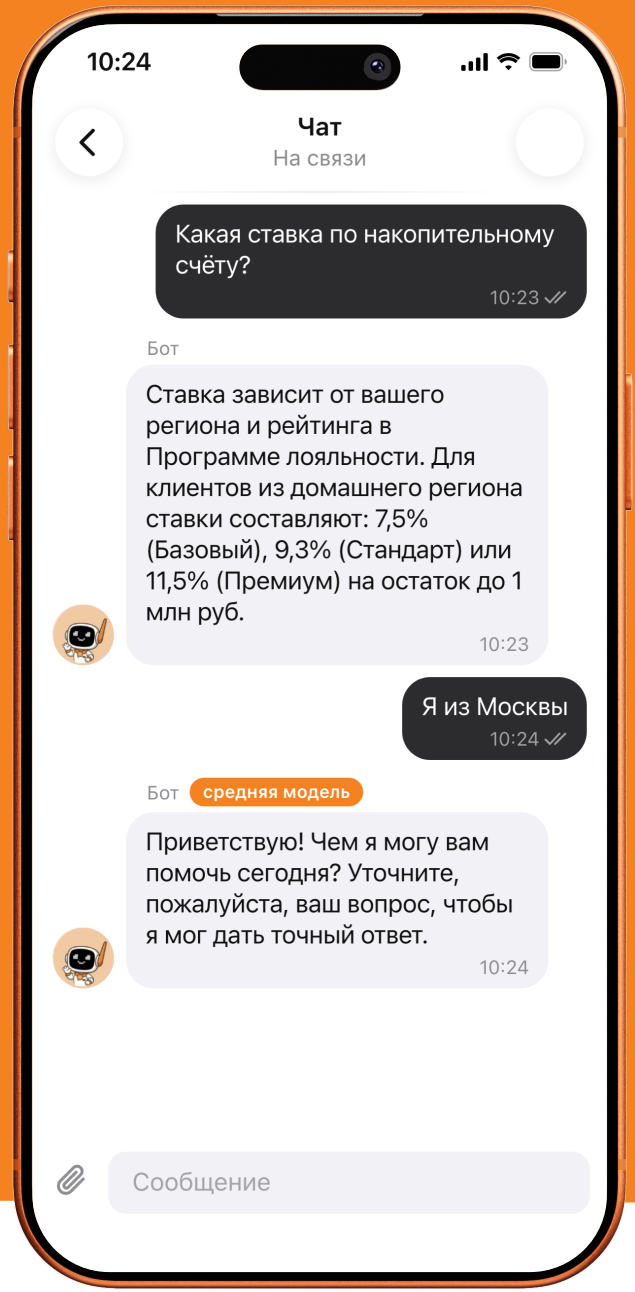
+24 п.п.
дала подготовка запроса без смены модели

8 из 8
запросов теперь решаются верно

Что выгоднее – купить модель втрое дороже или добавить слой подготовки запроса?



* 98 реальных вопросов клиентов банка



Без истории бот теряет нить

Что хотел узнать клиент

Ставку по накопительному счёту для Москвы

Что ушло в поиск

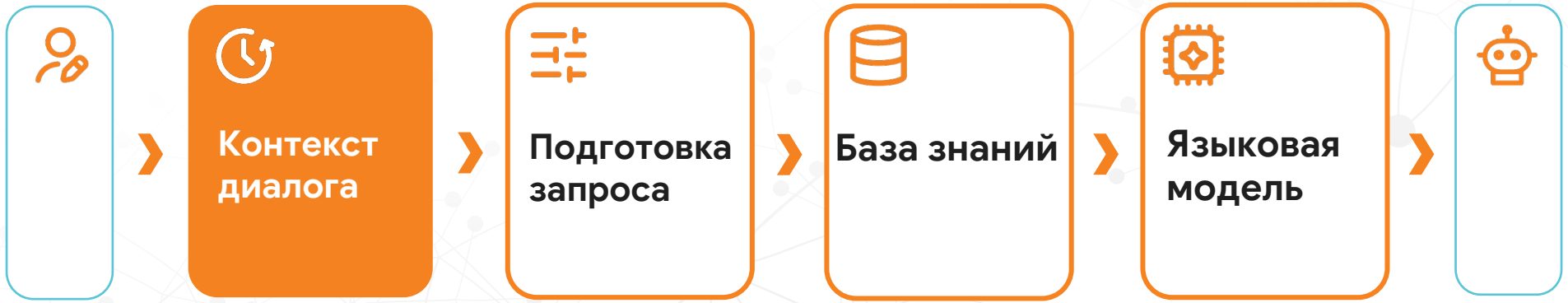
Только «я из Москвы» – ни слова про ставку

Что ответила средняя модель

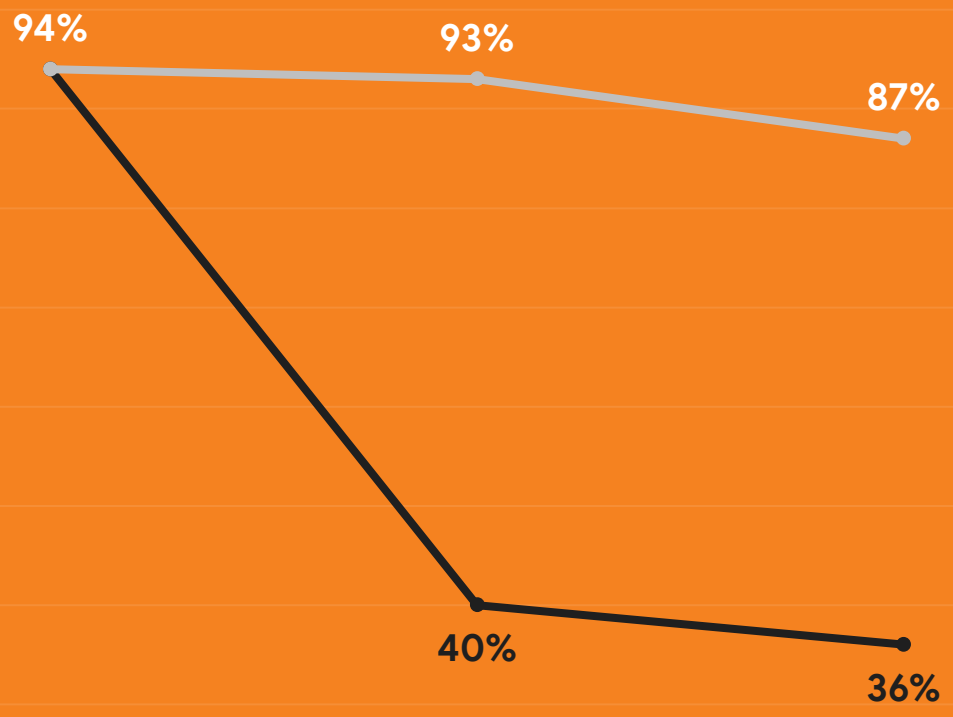
Поздоровалась и попросила уточнить вопрос

Модель не получила историю диалога.

Шаг 3: контекст диалога или как не потерять смысл на втором вопросе



Реплика клиента:	Что уходит в поиск:	Решение блока:
Я из Москвы	Какая ставка по накопительному счёту для клиентов из Москвы?	**дописать тему**
А сколько по времени?	Сколько по времени занимает перевыпуск карты?	**дописать тему**
Как заблокировать карту?	Как заблокировать карту? – без изменений	**новая тема**



Один и тот же бот – два разных результата

Всё решает контекст диалога.

57% в среднем по трём ходам без блока контекста

40% на втором ходу без блока контекста

93% на втором ходу с блоком контекста

Качество измеряется по номеру хода. Средняя цифра прячет провал.

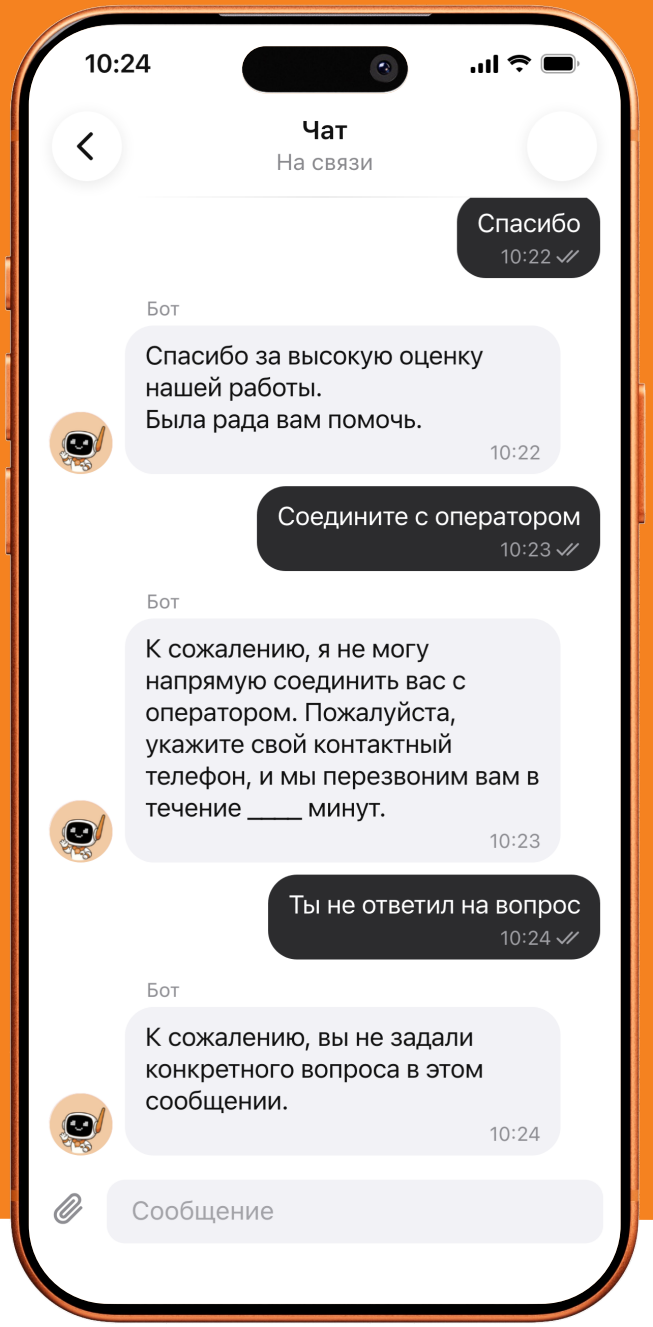
Не каждая реплика – вопрос к базе

Что пишет клиент?
«Спасибо», «Понял», «Соедини с оператором»

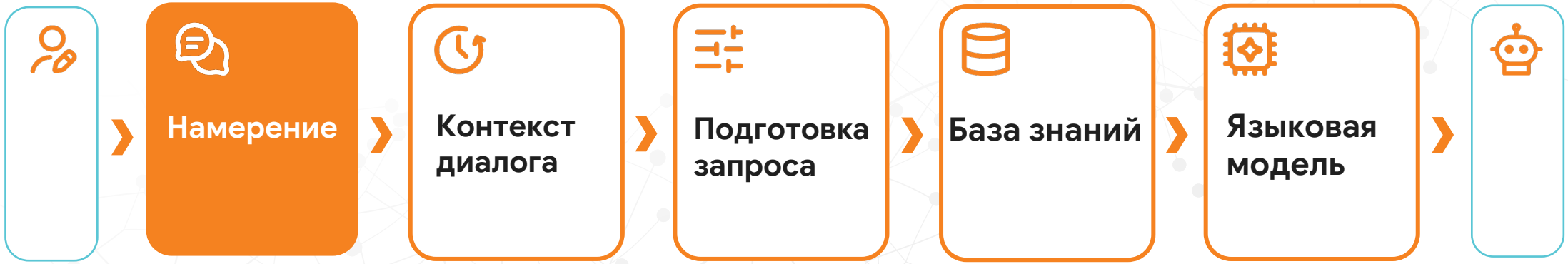
Что делает бот?
Ищет ответ в базе знаний, потому что не умеет различать намерения

Что получает клиент?
Шаблон, прочерк, раздражение

Бот не понимает, чего хочет клиент.



Шаг 4: определяем намерение клиента прежде, чем искать информацию



Результат:

Привет
поздороваться

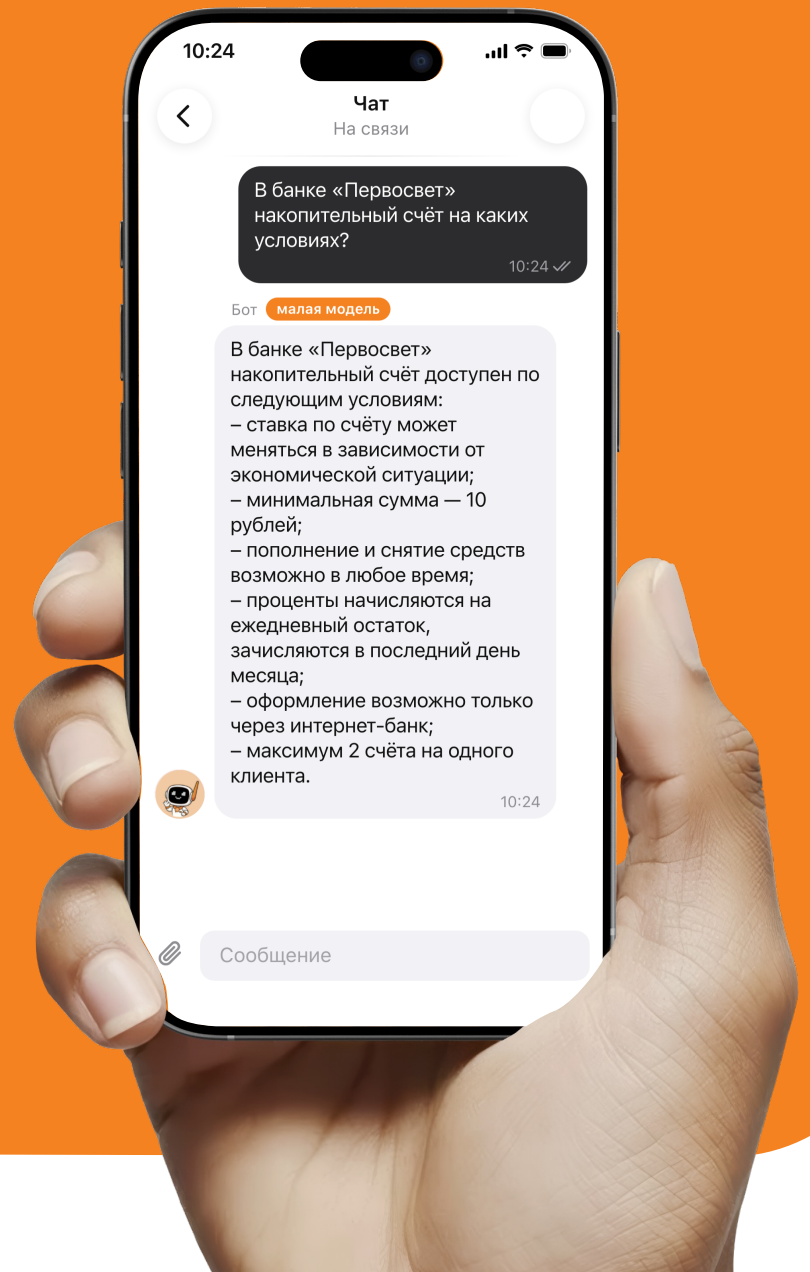
Спасибо
ответить без поиска

Понял
закрыть разговор

Уточнение
продолжить тему

Оператор
перевести на человека

Негатив
извиниться, позвать оператора



Бот выдал наши условия за чужие

Поиск

Статьи про накопительный счёт, релевантность 0,70

Проверки

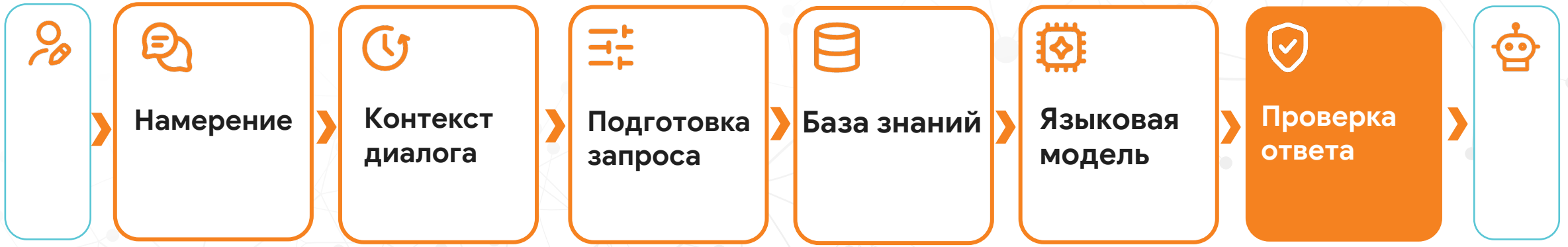
Пройдены: ответ по фрагментам базы

Итог

Ложные сведения о другом банке от имени нашего

Релевантность ≠ правильность.

Шаг 5: проверка ответа, чтобы ни одной ошибки не ушло клиенту



Цена проверки: **2** лишних отказа **+3** обращения к модели **+10 сек** на ответ

Честный отказ – дешевле ошибки.

Клиент судит по худшему ответу

Ради чего автоматизируем:

- ответ ночью
- ответ в выходной
- за секунды, без очереди

Оценка результата:

Мы – средним по всем клиентам

79% верных ответов в среднем

93% на втором ходу диалога

≠

Клиент – своим худшим ответом

1 ответ – тот, который не сработал

звонит в банк или уходит

Что спросить у подрядчика, чтобы не купить демо

Реальные обращения

Как бот отвечает на наших вопросах, а не на демо?

Метрики по ходам

Какое качество на втором и третьем ходу диалога?

Не только вопросы

Что бот делает со «спасибо» и просьбой об операторе?

Неверные ответы

Сколько их – отдельно от честного «не знаю»?

Проверка ответа

Что проверяется перед отправкой клиенту?

Работа с базой

Как база учит слова ваших клиентов?

iQSTORE
POWERED BY iCAM

Благодарим за внимание

Подписывайтесь на наши каналы

+7 (495) 120 31 82
info@iqstore.ru
iqstore.ru