

Почему большинство
генеративных голосовых
роботов ломаются при
внедрении

Арсений Назаркин, СТО



Почему вау эффект от
демо не работает в
реальности?

Ожидание



Реальность



Почему вау эффект от
демо не работает в
реальности?

Ожидание



Реальность



Почему **нельзя просто подключить** нейросеть к телефонии

МИФ (как это кажется пользователю)



Простой, мгновенный процесс

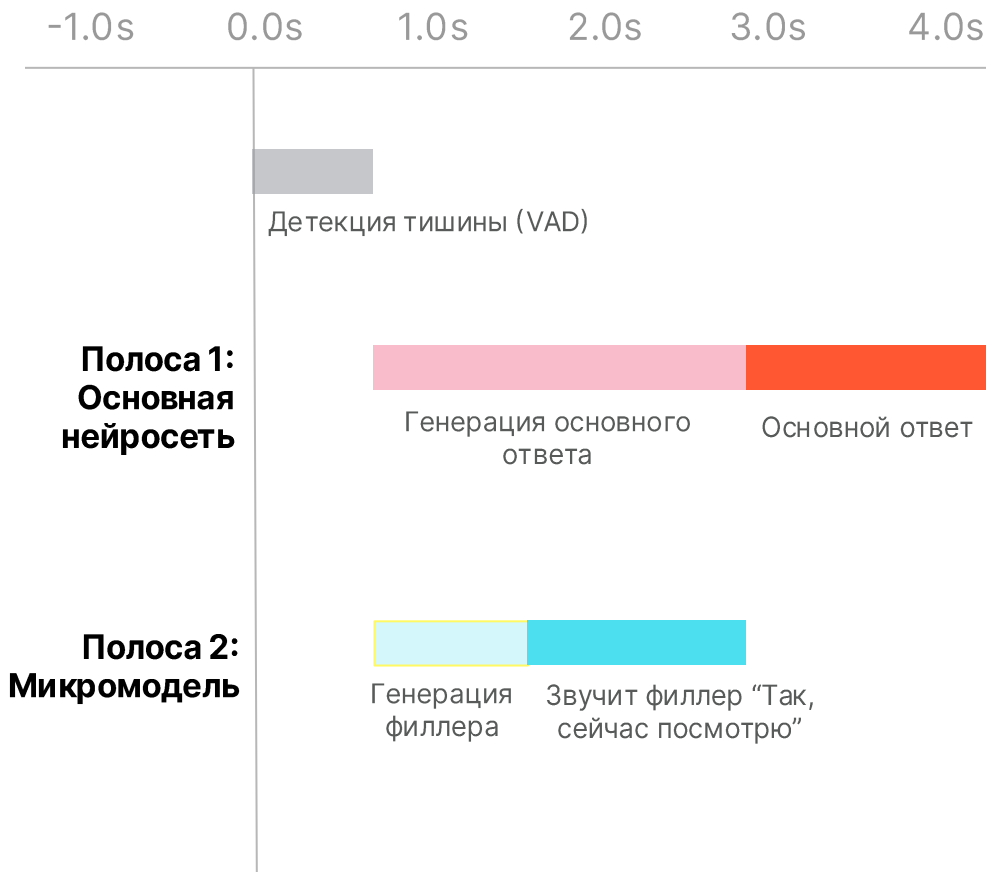
РЕАЛЬНОСТЬ (техническая архитектура)



ИТОГО: около 3 секунд

Техническая задержка процесса

Как скрыть задержку: магия контекстных филлеров



Почему **монолитные модели** ломаются о реальный бизнес

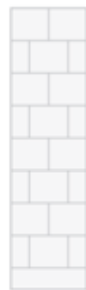
01 — МОНОЛИТНАЯ МОДЕЛЬ

End-to-End

Голос клиента



Persona Plex



Голос робота



GPT-Realtime / Moshi

Дорого и изолированно

02 — МОДУЛЬНАЯ АРХИТЕКТУРА

Оркестрация

Текстовый
запрос

Оркестратор

Ответ



Модульная система

Гибко и масштабируемо

Почему нельзя просто написать **один большой промпт**

01 — тяжёлый подход

Мега-промпт



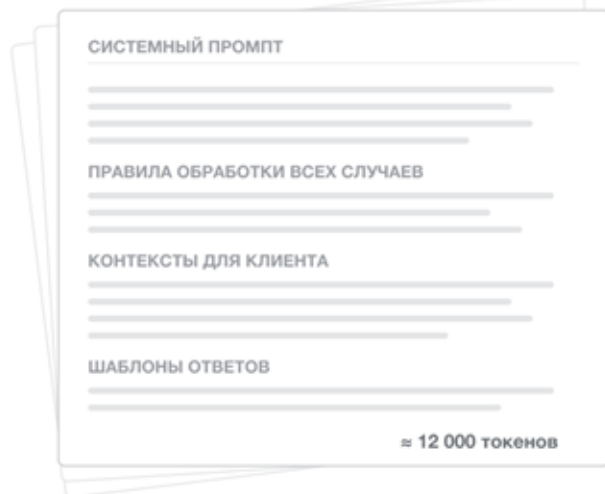
ДОРОГО



МЕДЛЕННО



ОШИБКИ



Один документ со всеми правилами загружается на каждый шаг диалога

02 — умный подход

Оркестратор



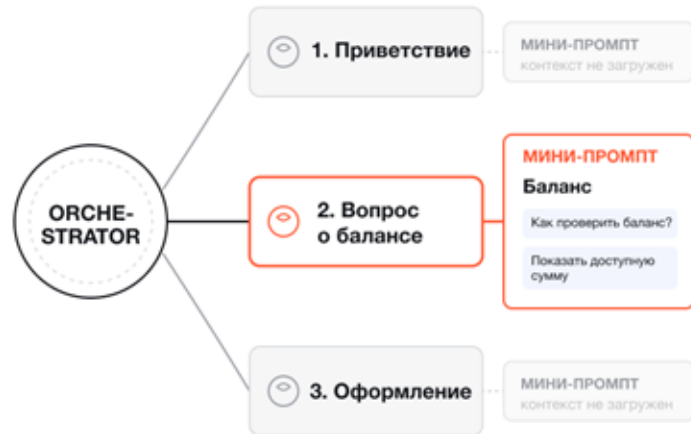
БЫСТРО



ТОЧНО



ЭКОНОМНО



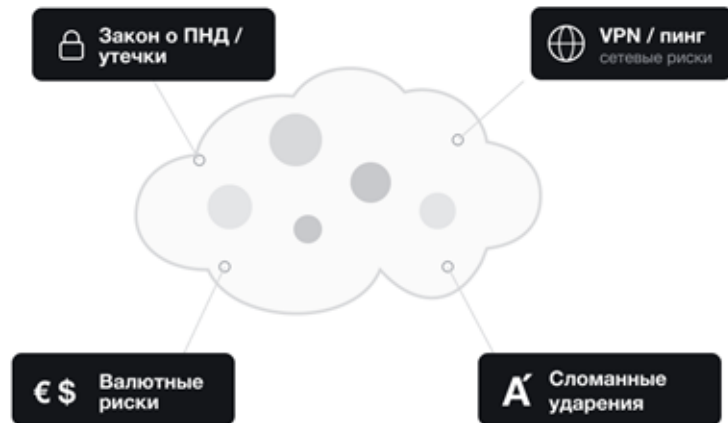
Оркестратор собирает контекст под конкретный шаг диалога

Ловушка западных технологий и наше решение

01 — ВНЕШНИЕ СЕРВИСЫ

Риски

🔊 Cartesia



Зависимость от внешних провайдеров — юридические, сетевые и языковые риски

02 — РЕФЕРЕНСНЫЙ СИНТЕЗ

Эlegantное решение

БЛОК 1 — ТЕКСТ

«Вы подтверждаете оплату?»

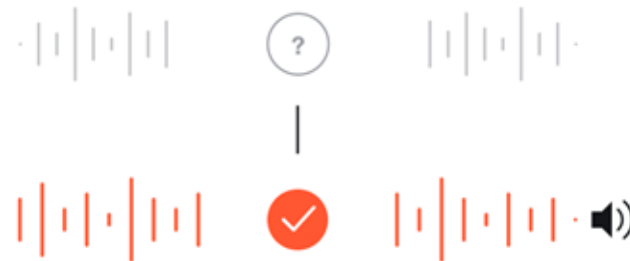


🔊 Яндекс 1

🔊 Яндекс 2

🔊 КВИНТ

БЛОК 2 — + АУДИО-РЕФЕРЕНС



ИДЕАЛЬНЫЙ СИНТЕЗ С ВОПРОСИТЕЛЬНОЙ ИНТОНАЦИЕЙ

Текст задаёт смысл, аудио-референс — интонацию

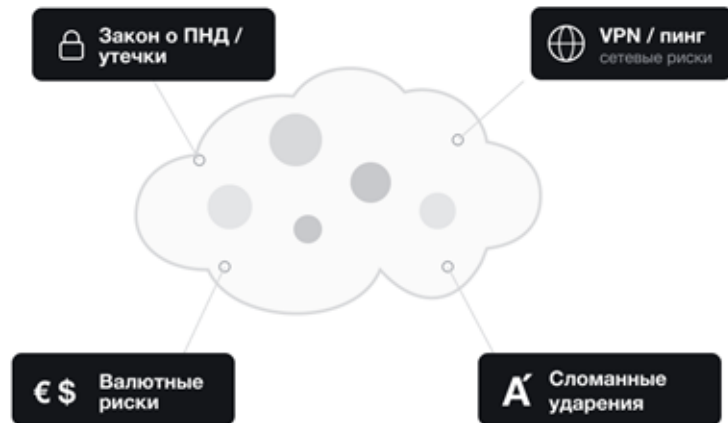
Текст и аудио-референс — точная интонация без внешних сервисов

Ловушка западных технологий и наше решение

01 — ВНЕШНИЕ СЕРВИСЫ

Риски

🔊 Cartesia



Зависимость от внешних провайдеров — юридические, сетевые и языковые риски

02 — РЕФЕРЕНСНЫЙ СИНТЕЗ

Эlegantное решение

БЛОК 1 — ТЕКСТ

«Вы подтверждаете оплату?»

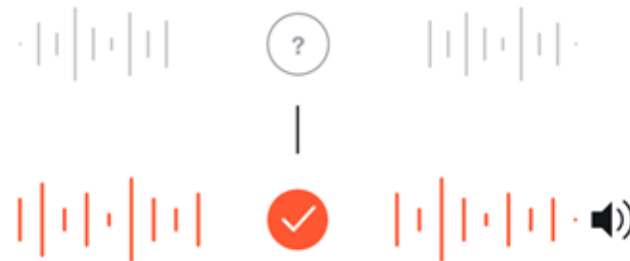


🔊 Яндекс 1

🔊 Яндекс 2

🔊 КВИНТ

БЛОК 2 — + АУДИО-РЕФЕРЕНС



ИДЕАЛЬНЫЙ СИНТЕЗ С ВОПРОСИТЕЛЬНОЙ ИНТОНАЦИЕЙ

Текст задаёт смысл, аудио-референс — интонацию

Текст и аудио-референс — точная интонация без внешних сервисов

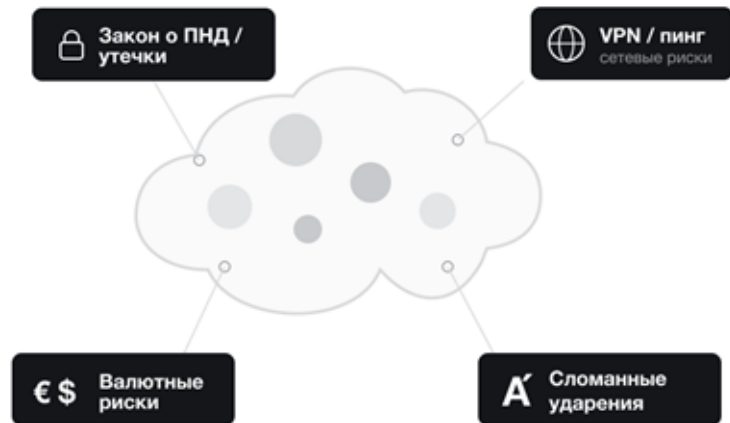
Ловушка западных технологий и наше решение

01 — ВНЕШНИЕ СЕРВИСЫ

Риски



Cartesia



Зависимость от внешних провайдеров — юридические, сетевые и языковые риски

02 — РЕФЕРЕНСНЫЙ СИНТЕЗ

Эlegantное решение



Яндекс 1



Яндекс 2



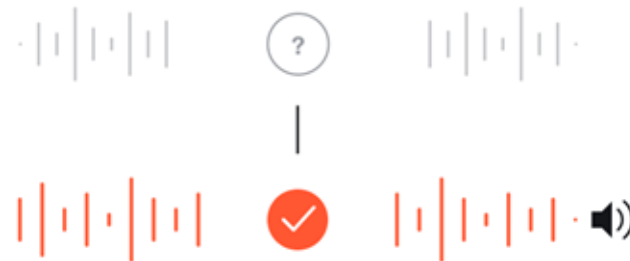
КВИНТ

БЛОК 1 — ТЕКСТ

«Вы подтверждаете оплату?»



БЛОК 2 — + АУДИО-РЕФЕРЕНС



**ИДЕАЛЬНЫЙ СИНТЕЗ С ВОПРОСИТЕЛЬНОЙ
ИНТОНАЦИЕЙ**

Текст задаёт смысл, аудио-референс — интонацию

Текст и аудио-референс — точная интонация без внешних сервисов

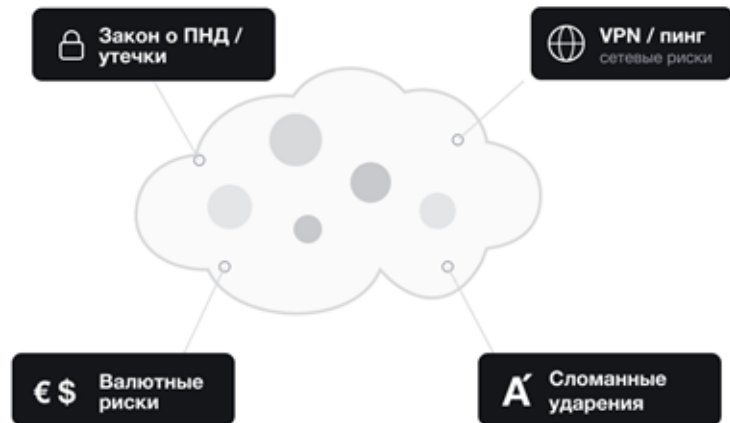
Ловушка западных технологий и наше решение

01 — ВНЕШНИЕ СЕРВИСЫ

Риски



Cartesia



Зависимость от внешних провайдеров — юридические, сетевые и языковые риски

02 — РЕФЕРЕНСНЫЙ СИНТЕЗ

Эlegantное решение



Яндекс 1



Яндекс 2



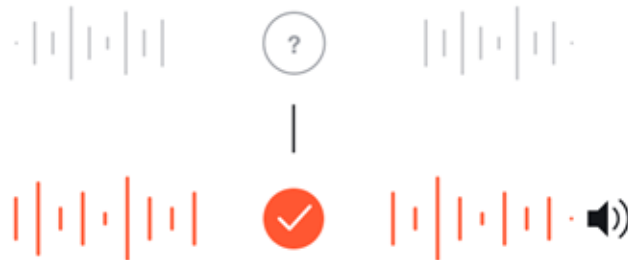
КВИНТ

БЛОК 1 — ТЕКСТ

«Вы подтверждаете оплату?»



БЛОК 2 — + АУДИО-РЕФЕРЕНС

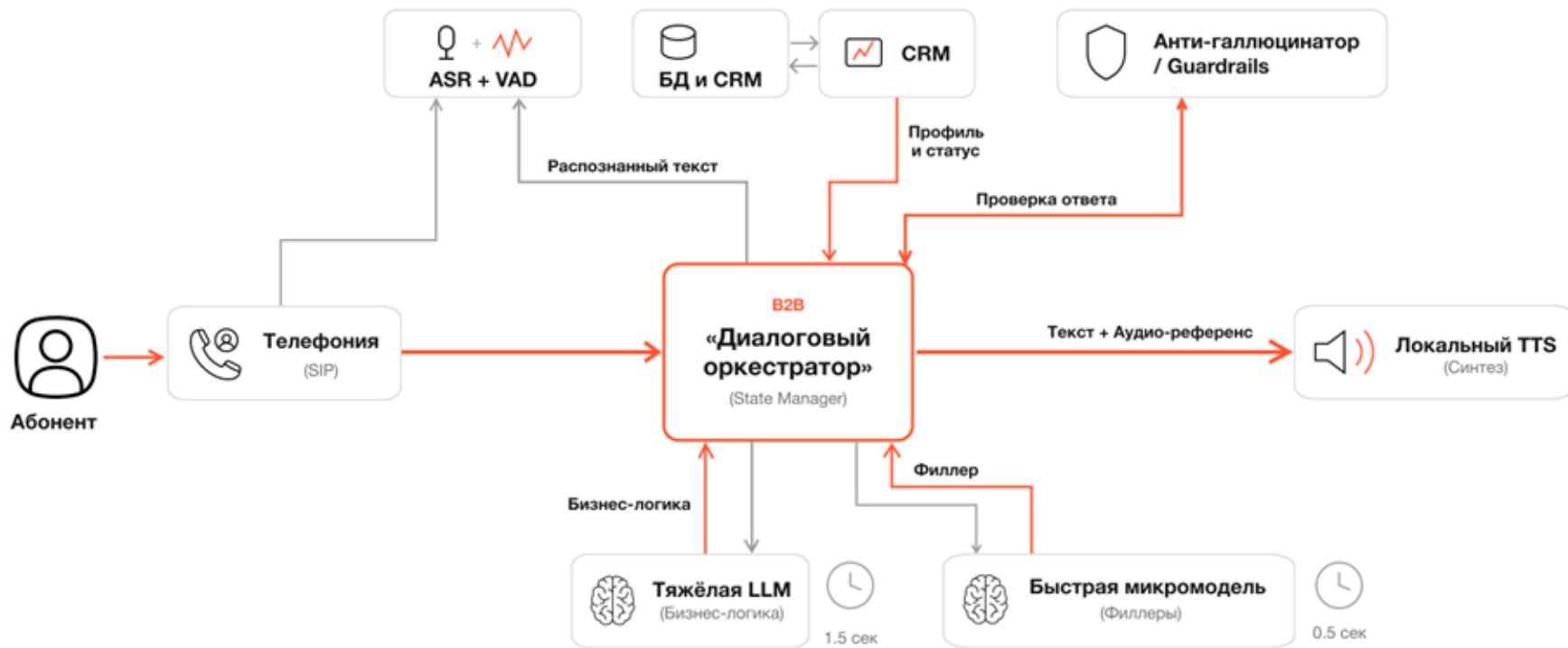


**ИДЕАЛЬНЫЙ СИНТЕЗ С ВОПРОСИТЕЛЬНОЙ
ИНТОНАЦИЕЙ**

Текст задаёт смысл, аудио-референс — интонацию

Текст и аудио-референс — точная интонация без внешних сервисов

Анатомия генеративного головного робота





Арсений Назаркин
СТО

arseny.nazarkin@kvint.io
www.kvint.io

