

Голосовой ИИ-агент в платформе Эра

Вместе с операторами, а не вместо них



методология • архитектура • инструменты



Содержание

- СОРС 8.0: что изменилось в стандарте
- Итерационный цикл: цель — метрика — эксперимент — правка
- Технологии: что такое ASR, TTS, LLM, VAD
- ИИ-робот: как добавить голосового агента в очередь
- ИИ-помощник: сбор контекста, пока клиент ждёт на линии
- ИИ-суфлёр: подсказки оператору во время работы
- Эволюция подходов: от жёстких сценариев к гибким моделям
- Новое в текстовых каналах
- Развертывание: облачные сервисы или закрытый контур
- Производительность и сайзинг



А ТЫ УЖЕ ПРИМЕНЯЕШЬ СТАНДАРТ COPC 8.0?



КЛИЕНТ В ЦЕНТРЕ



ЭФФЕКТИВНОСТЬ



КАЧЕСТВО



УЛУЧШЕНИЕ



РАЗВИТИЕ

— УПРАВЛЯЙ ПО ЛУЧШИМ ПРАКТИКАМ. —
ДОСТИГАЙ БОЛЬШЕГО! —

COPC 8.0: стандарт управления КЦ

COPC Customer Experience (CX) Standard for Contact Centers

Suitable for Business Process Outsourcers (BPOs) and
Customer Service Providers (CSPs)

RELEASE 8.0
VERSION 1.0

copc



СОРС: основные тезисы

- Автоматизация и ИИ входят в ключевые факторы управления КЦ
- ИИ перестал быть инструментом, а стал средой работы
- Этика и ответственность применения технологий, особенно ИИ
- Качество: оценка, калибровка, отключение
- Люди и технологии, выполняющие клиентские задачи



СОРС: управление технологией

- Точки влияния технологии на клиентский опыт
- Проверка корректности работы технологии
- Калибровка
- Границы допустимого применения
- Ответственность за последствия



Итерационный цикл



Цель • Метрика • Эксперимент • Правка



Демо

Страховая компания



Компоненты голосовых роботов

LLM

ASR

TTS

VAD

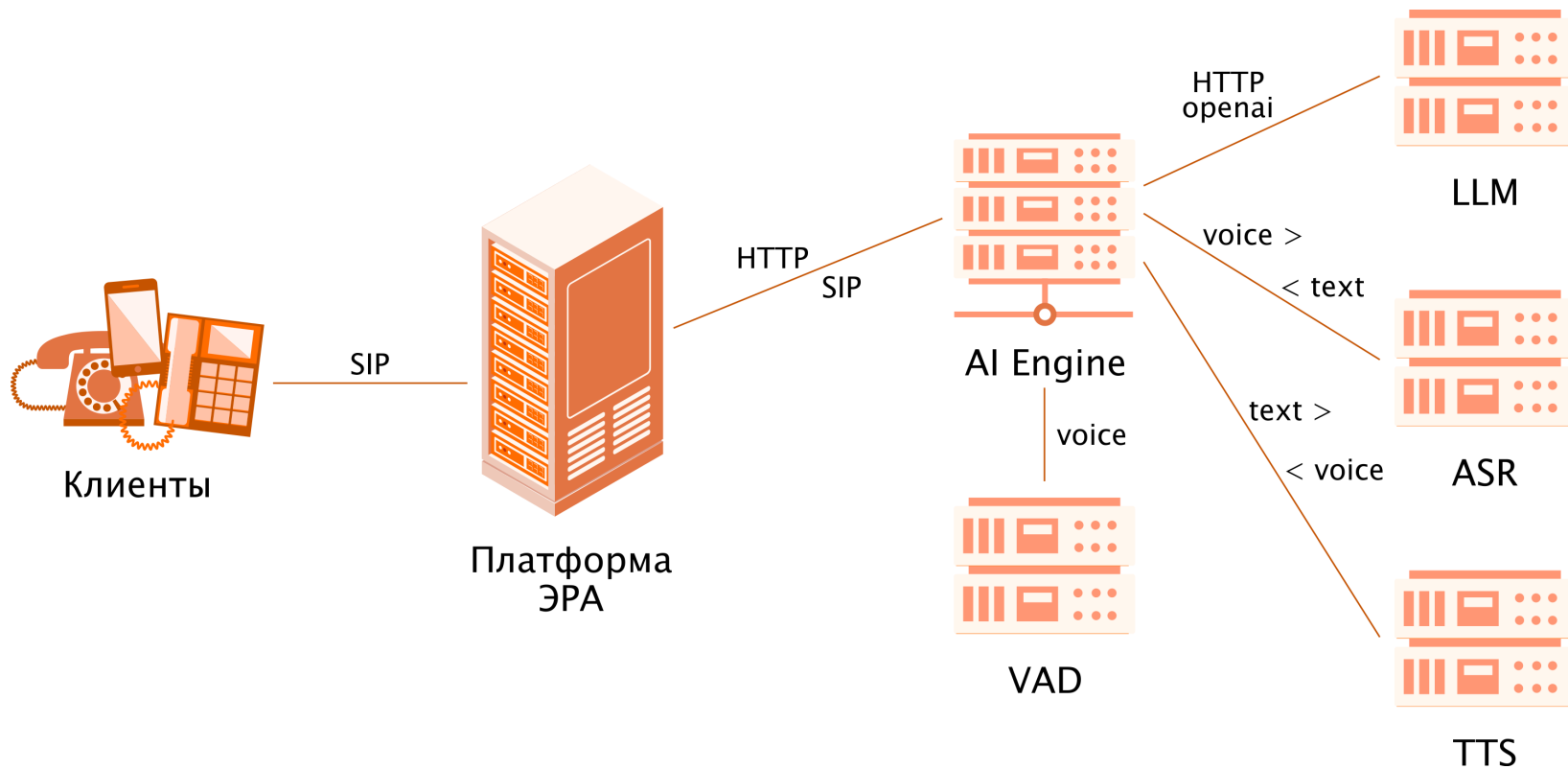


Context

Tools



Архитектура голосовых роботов



Демо: ИИ-робот

Как добавить голосового агента в очередь

Урегулирование задолженности

Согласование доставки



Отличительные особенности

- Реагирует на перебивания
- Понимает смысл: да нет наверное
- Предлагает четкие шаги и фиксирует обещание
- Конвертирует даты и числительные
- Понимает: повторите, ничего не поняла
- Отрабатывает предложение отложить звонок
- Пытается проявить эмпатию
- Неустанно работает в сложных условиях



ИИ-помощник: сбор контекста

- Включается в очереди ожидания
- Оставляет возможность отказаться
- Собирает информацию в свободной форме
- Упорядочивает и структурирует
- Отображает оператору на карточке



ИИ-суфлёр: подсказки оператору

- Слушает разговор или диалог
- Анализирует контекст (сторону клиента и/или оператора)
- Отображает статьи базы знаний
- Предлагает заготовки ответов
- Фрагменты стенограммы для карточки





Эволюция подходов

- Классический сценарий
- Сценарий с элементами ИИ
- Единый промпт-инструкция
- Агентский подход: четкие этапы и цели



Развертывание

- Независимость компонентов: LLM, ASR, TTS
- Универсальность протокола LLM (openai)
- Разнородные протоколы ASR, TTS (коннекторы)
- Информационная безопасность и персональные данные
- Риски обновления версий и недоступности сервисов
- Минимизация сетевых задержек
- Приемлемые требования к инфраструктуре



Компоненты и расход видеопамяти

Компонент	Модель	Общий объем	Объем сессии
LLM	Qwen3.5-35b-a3b-awq	23 GB	200 MB
ASR	T-one 71M	300 MB	100 MB
TTS	FunAudio CosyVoice 3	8 GB	200 MB



Производительность видеокарт NVIDIA

Модель	Память	Вычисления	Количество сессий
RTX 5090	32 GB	105 TFLOPS	1-3
RTX PRO 5000	48 GB	67 TFLOPS	20-30
RTX PRO 6000	96 GB	126 TFLOPS	40-60





Заключение



Платформа ЭРА

- Омниканальный контакт-центр
- Распределенная IP-АТС
- Открытая конфигурация и Builder
- Микросервисная архитектура
- Распределенные кластеры
- Мультидоменность
- Облака и закрытые контуры



**ТОВАРИЩИ,
НАНИМАЙТЕ НАШИХ
ИИ-ОПЕРАТОРОВ**



**== ДЛЯ УСИЛЕНИЯ ==
ВАШЕГО КОНТАКТ-ЦЕНТРА!**

Спасибо!

