

1 — Опорный стек: один SFU + паттерн агента

Медиа-плоскость	Клиенты -> SFU -> клиенты. WebRTC + Opus. Пересылает, не перекодирует.
AI-плоскость	Один агент в комнате как участник: STT -> субтитры, перевод, заметки.
Края + контроль	Сервер токенов (auth) - TURN/coturn (NAT) - бэкэнд (транскрипты).

2 — Build vs Buy (2026): берите инфру, покупайте модели, стройте продукт

ВЗЯТЬ SFU (LiveKit/mediasoup) · фреймворк агентов (LiveKit Agents) · TURN (coturn)
КУПИТЬ распознавание речи (Deepgram/AssemblyAI) · перевод (frontier-LLM/MT API)
СТРОИТЬ on-device-размытие (MediaPipe+WebGPU) · сервер токенов · бэкэнд · UX
ВНИМАНИЕ Krisp платный с 2026-05-01; RNNoise без поддержки с 2024; DeepFilterNet free.

3 — Два параллельных бюджета (с арифметикой)

ИНТЕРАКТ. МЕДИА	~200 ms в одну сторону. Если сеть = 120 ms, остаётся ~80 ms на захват + on-device-AI + кодирование + воспроизведение.
AI-ГОЛОС. ЦИКЛ	~800 ms. VAD 50 + STT 150 + LLM 400 + TTS 150 + сеть 50 = 800 ms. Потоково каждый этап; обученный детектор конца реплики, не таймер.

4 — Уровни масштабирования и правило стоимости

<100 -> один узел SFU. 100-1K -> simulcast + SVC (500-800 видео/узел).
1K+ -> каскадные SFU по регионам (ретранслятор = ещё один участник).
СТОИМОСТЬ on-device = \$0 · SFU ~\$0,0005/участник-мин · агент ~\$0,01/мин.
Час на 25 человек, 5 функций: ~\$0 + ~\$0,75 + ~\$1,20 = < \$2,00.
Self-host SFU после ~500 устойчивых одновременных сессий.

5 — Порядок сборки: каждый майлстоун — рабочий продукт

1 Простой звонок (SFU+токен+TURN) 2 On-device размытие+шум 3 Агент субтитров (STT)
4 Перевод + заметки (тот же агент) 5 Модерация/аватары + уведомление по AI Act.

6 — Чек-лист запуска (инженерный контекст, не юр. консультация)

- ☐ Сначала твёрдый простой звонок: звук/видео + переподключение на реальных устройствах.
- ☐ AI на человека (размытие, шум) — на устройстве, никогда в облаке на каждого участника.
- ☐ Ровно ОДИН шумоподаватель на сыром звуке до Opus; выключите встроенный, если добавили свой.
- ☐ SFU не перекодирует; модерация на копии-ответвлении, не внутри медиасервера.
- ☐ TURN с ограниченными по времени учётными данными; ~15-20% сетей без него не соединятся.
- ☐ Наблюдаемость на сессию (время связи, потеря пакетов, задержка агента) с майлстоуна 1.
- ☐ Раскрытие AI-сгенерированного голоса/лица — EU AI Act ст. 50 (с 2026-08-02).