

## 1 — Что такое живой субтитр

Текст с таймингом и говорящим (cue WebVTT: текст + говорящий + начало/конец).  
Создаёт потоковый ASR; показ через ~1–2 с после произнесённых слов.

## 2 — Паттерн fan-out ASR на стороне SFU

Звук говорящего -> SFU (роутер, что вы и так держите) -> VAD-гейт -> ОДИН ASR.  
Субтитровый cue -> fan-out по data channel -> все видят одну и ту же строку.  
Два fan-out на одном сервере: звук — чтобы слышать, текст — чтобы читать.

## 3 — Стоимость (часовая встреча на 30 человек)

Наивно (каждая дорожка):  $30 \times 60 \text{ мин} \times \$0,0077/\text{мин} = \$13,86$  за встречу.  
Fan-out + VAD (~1,5 говорящих):  $1,5 \times 60 \times \$0,0077 = \$0,69$  за встречу.  
~20x дешевле. Рычаг — VAD-гейтинг: распознавать только активную речь.

## 4 — Черновик против финала

Черновик (`is_final:false`) = быстрая догадка, изменится. Финал (`is_final:true`) = фикс.  
Правило клиента: ЗАМЕНЯТЬ незавершённую строку до финала — не добавлять черновики.

## 5 — Доставка и черта доступности

Доставка текста по data channel WebRTC (RFC 8831) — надёжно + упорядоченно.  
Живые субтитры = WCAG 1.2.4 уровень AA. Правило DOJ ADA: WCAG 2.1 AA обязателен.  
Дедлайн: 24 апр 2026 (госорганы >50k); 26 апр 2027 (меньшие).

## 6 — Чек-лист «строить или покупать»

- ☐ Групповой звонок, который вы записываете / переводите? -> серверный fan-out на SFU.
- ☐ 1:1 или офлайн / приватно, звук не должен покидать устройство? -> клиентский ASR.
- ☐ Self-host медиасервера? -> mediasoup / Janus / Jitsi (свой аудио-съём).
- ☐ Хотите субтитры почти-настройкой, не проектом? -> LiveKit (встроенный поток).
- ☐ Выбирайте ASR по задержке до 1-го черновика + точности на шуме + цене.
- ☐ Всегда VAD-гейтите распознавание — платите за говорящих, не за места.