Памятка: выбор архитектуры перевода в реальном времени

Каскад против end-to-end, окно 2-3 секунды, WebRTC sidecar и сравнение систем 2026 года

1 · Четыре системы 2026 года с одного взгляда

Система	Что это	Как получить	Языки	Когда лучше
Google Meet	Встроенная фича	Подписка	EN <-> 5 языков	Обычные встречи; голос как ваш
OpenAI translate	Модель в аренду	API (~$0,034/мин)	70 вход / 13 выход	Строить продукт; динамич. голос
SeamlessStreaming	Открытая модель	Свой хостинг	~100 вход. языков	Открытый эталон; CC-BY-NC (неком.)
Корп. синхрон	Сервис по контракту	Контракт	6000+ комбинаций	Ответств. события; человек в контуре

Google Meet, OpenAI и Seamless — end-to-end-движки; корпоративные платформы (KUDO, Interprefy, Wordly) держат живого синхрониста для точности, которую один AI пока не гарантирует. Число языков и цены прочитаны в мае 2026 — перепроверьте при использовании.

2 · Каскад или end-to-end?

Критерий	Каскад (3 модели)	End-to-end (1 модель)
Как работает	Связка ASR -> MT -> TTS, читаемый текст между этапами	Одна модель: вход язык A, выход язык B
Задержка	~1-2 с; растёт за 3 передачи	~400-700 мс; меньше передач
Отлаживаемость	Высокая - читаем расшифровку на каждом этапе	Низкая - нет читаемой середины
Голос	Уплощён в текст и обратно	Хранит тон, просодию, голос говорящего
Привязка	Меняй любой компонент, любого вендора	Один вендор на все три
Когда брать	Аудиторский след, регуляция, отладка	Важнее низкая задержка и естеств. голос

3 · Перед запуском фичи перевода в реальном времени

- [ ] Назван формат: устный перевод (голос) или субтитры, или и то и другое на выбор пользователя?
- [ ] Выбрана архитектура: каскад для контроля/отладки, end-to-end для низкой задержки и естественного голоса.
- [ ] Задана цель по задержке в окне 2-3 секунды - быстрее трудно понять, медленнее рушит разговор.
- [ ] Решена точка скорость-точность: ждать дольше ради точности или фиксировать быстрее ради плавности - не оба.
- [ ] Подтверждено, что задержка держится ниже ~2 секунд в диалоге, иначе участники начнут говорить поверх друг друга.
- [ ] Выбран один из четырёх типов систем: Google Meet, OpenAI gpt-realtime-translate, SeamlessStreaming или корп. синхрон.
- [ ] Проверен охват языков для нужных пар, включая структурно трудные (например, немецкий с глаголом в конце).
- [ ] Спланирован транспорт: sidecar снимает аудио SFU, переводит и подаёт дорожки по языкам слушателям обратно.
- [ ] Заложена одна сессия на язык, а не на участника (3 языка = 3 сессии, которые SFU разводит).
- [ ] Исключён дословный перевод: выбрана модель, переводящая смысл, чтобы идиомы и сарказм не стали бессмыслицей.
- [ ] Получено согласие на клонирование голоса участника - подключение к звонку не является согласием (NO FAKES Act).
- [ ] Добавлены происхождение и безопасность: водяной знак на сгенерированном аудио и детекция добавленной токсичности.
- [ ] Сохранён человеческий откат для ответственного контента (правление, звонки инвесторам, клинические консультации).