

1 - Три схемы (выбор по тому, что значит видео)

Transcript-first . только субтитры . <\$0,01 / 30-мин . лекции, подкасты, речь.
Гибрид транскрипт + кадры . ~\$0,05-0,25 . вебинары, tutorиалы, screen-share.
Мультимод. смотрит кадры + звук . ~\$0,25-1,35 . демо, спорт, видеонаблюдение.

Решающий вопрос: смысл живёт в СЛОВАХ или в КАРТИНКЕ?

2 - Пайплайн из пяти этапов (общий для всех схем)

- 1 Достать слова .. готовые субтитры (free) ИЛИ распознавание речи (Whisper ~10% WER).
- 2 Почистить/резать сшить реплики в предложения; чанки (~10k символов / 1k перекрытие).
- 3 Суммаризировать stuff / map-reduce / refine (см. блок 3).
- 4 Оформить вывод . тезисы, главы, JSON; тайм-коды (MM:SS).
- 5 Проверить барьер точности - сверить каждое утверждение с источником.

3 - Суммаризация длинного транскрипта + арифметика стоимости

Map-reduce чанки параллельно, затем объединить - быстро; не связывает далёкие части.
Refine текущая выжимка вперёд - рассуждает по всему; последовательно, медленнее.
Stuffing ... весь транскрипт в один long-context промпт - вариант по умолчанию.

30-мин видео: транскрипт ~6 000 токенов = <\$0,01 | мультимод. 1 FPS ~300 ток/сек
-> 1 800с x 300 = 540 000 токенов ~ \$1,35 (полное) | низкое ~100 ток/сек ~ \$0,23.

4 - Своё или купить (аренду модель, владе пайплайном)

Аренда API . frontier-модель + hosted STT - быстро, за запрос; низкий/скачкообразный объём.
Своё Whisper + open LLM - фикс. цена; высокий стабильный объём или приватность.

5 - Барьер выпуска - качество + право (контекст, не юр. консультация)

- ☐ Выбирайте схему по контенту, а не по привычке: кадры - только когда смысл визуален.
- ☐ Привязывайте каждый тезис и тайм-код к строке транскрипта - пусть пользователь кликнет и проверит.
- ☐ Оценивайте точность через LLM-as-judge, а не пересечение слов (ROUGE) - галлюцинация опаснее всего.
- ☐ Кэшируйте длинные видео, которые переспрашиваете; считайте точку аренда-vs-своё до старта.
- ☐ Доступ к субтитрам только через санкционированные пути API - бесплатные скрейперы часто запрещают коммерцию.