

Ролик, который монтирует себя сам

Как устроена система, которая берёт тему и отдаёт готовое видео: озвучка, субтитры по словам, переходы, запись экрана, липсинк и публикация. Без волшебной кнопки — но и без ручного монтажа.

Для кого. Ты снимаешь ролики для соцсетей, YouTube или обучения и устал монтировать руками. Опыт программирования не нужен: всё, что ниже, — это описание ролей и порядка, а команды выполняет помощник.

- | | | | |
|---|--------------------------------|---|--------------------------------|
| 1 | Кто за что отвечает | 2 | Что подготовить один раз |
| 3 | Порядок сборки: 12 шагов | 4 | Правила, которые экономят часы |
| 5 | Звук: три эффекта из воздуха | 6 | Чеклист перед публикацией |
| 7 | Длинные ролики: 30 минут и час | 8 | Если не разобрался |

01 Кто за что отвечает

Главное непонимание новичка: он ищет одну нейросеть, которая «сделает видео». Такой нет. Есть бригада, где у каждого одна работа, а помощник — прораб, который их вызывает по очереди.

КТО	ЧТО ДЕЛАЕТ	ЧЕМ ЗАМЕНИТЬ
Claude Code (или другой ИИ-помощник в терминале)	прораб: читает твои инструкции, вызывает остальных, собирает результат	Codex, Cursor
Whisper	слушает речь и отдаёт текст с таймингом каждого слова — из этого получаются субтитры	любой сервис распознавания с потайминговкой
FFmpeg	рабочая лошадь: режет, склеивает, накладывает картинку на картинку, сводит звук	ничем, это стандарт
Сервис озвучки (например, Fish Audio)	читает текст голосом — своим клонированным или готовым	ElevenLabs и аналоги
Playwright	открывает сайт и снимает экран — то самое «видео сверху»	запись экрана руками
Скрипты на Python	мелкая работа: маски, обложки, перевод распознавания в файл субтитров	—

Вывод, ради которого написан гайд. Ты не учишь нейросеть монтировать. Ты один раз описываешь, кто из бригады что делает и в каком порядке, — и дальше помощник повторяет это на каждом ролике.

02 Что подготовить один раз

Это и есть та самая «система», без которой выходит разнобой. Собирается за вечер, живёт годами.

ЧТО	ЗАЧЕМ
Инструкции для помощника (у меня их четыре)	правила сценария, сборки, звука и публикации — чтобы не объяснять заново каждый раз
Пак звуков	одинаковые переходы во всех роликах
Шрифты в отдельной папке	субтитры и плашки не разъезжаются по стилю
Фон-подложка (короткое зацикленное видео)	кадр не выглядит пустым
Библиотека логотипов сервисов	вставлять логотип, о котором речь, без поиска каждый раз
Музыка без прав	чтобы площадка не приглушила звук и не сняла монетизацию

Про исходники

Нужны три вещи: клипы с говорящим (снятые или сгенерированные), запись экрана с процессом и свежий скриншот интерфейса. Всё вертикальное — 1080×1920, 24 кадра в секунду.

03 Порядок сборки: 12 шагов

Так собирается один ролик. Каждый шаг делает помощник, от тебя — тема и проверка глазами в конце.

1. Распознавание речи

Whisper слушает каждый клип и отдаёт файл со словами и точным временем каждого слова. Именно потайминговка даёт субтитры, где слово подсвечивается ровно тогда, когда оно звучит.

2. Сшивка таймингов

Клипы идут подряд, поэтому времени второго клипа прибавляется длительность первого, и так далее. Мелочь, которая ломает всё: на стыках слова слипаются («идеюне»), если не добавить пробел в начало каждого следующего куска.

3. Субтитры

Из распознавания собирается файл субтитров. Дальше — косметика: кодовое слово заглавными, названия сервисов с большой буквы.

4. Склейка речи

Клипы соединяются в одну дорожку. Проверь длительность: она должна совпасть с суммой кусков секунда в секунду.

5. Маска говорящего

Спикер не вырезается, а показывается через мягкую маску — скруглённый прямоугольник плюс овал под лицо. Размывать края нужно отдельно, иначе замылится либо рамка, либо лицо.

6. Основа кадра

Тёмный фон, поверх — подложка, поверх — спикер по маске. Фон растягивается на весь кадр: если обрезать по бокам, будет виден стык.

7. Карточки интерфейса

Запись экрана вставляется целиком, без обрезки по краям — иначе зритель видит обрывок и не понимает, что происходит. Лёгкий наезд делается плавным увеличением, а не встроенным «зумом» — тот дёргает картинку.

8. Появление карточек

Карточка выезжает снизу за полсекунды. За полсекунды до неё — искра-переход, она прикрывает стык и делает склейку незаметной.

9. Плашки с текстом

Текст на экране кладётся отдельным файлом настроек, а не строкой команды: иначе кавычки и переносы ломают сборку. Многострочную плашку собирай двумя строками, а не переносом внутри одной.

10. Звук

Три эффекта генерируются на месте, скачивать нечего (см. раздел 5). Дальше всё сводится с музыкой и речью в один слой с ограничителем громкости, чтобы не было перегруза.

11. Финальный проход

Субтитры и сведённый звук накладываются одним проходом — так быстрее и без потери качества от лишних пересжатий.

12. Обложка

Обложка — это кадр, который клеится в самое начало на треть секунды. Название сервиса или тема обязаны быть в первой строке, иначе в ленте непонятно, о чём ролик.

04 Правила, которые экономят часы

Это то, что выясняется на своей шкуре. Отдаю сразу.

- ☐ **Рабочая папка — только латиницей.** Кириллица в пути ломает наложение субтитров. Копируй исходники во временную папку с латинскими именами.
- ☐ **Проверяй распознавание глазами.** В тишине в конце клипа модель дописывает несуществующие фразы вроде «Субтитры делал...». Всё, что появилось после конца речи, — выбрасывать.
- ☐ **Мусорные слоги** («mmm», «ком») на месте пауз означают, что клип сгенерировался с тире в промпте. Проще пересоздать клип, чем чинить звук.
- ☐ **Никакого «зума» встроенным инструментом.** Он дрожит на длинных наездах. Плавное увеличение + фиксированная обрезка дают чистое движение.
- ☐ **Проверяй раскадровкой, а не превью.** Превью врёт: замыленный край маски и слипшиеся субтитры видны только на реальных кадрах.
- ☐ **Держи вес меньше 50 МБ**, если публикуешь в Telegram, — иначе он пережмёт сам и хуже.

05 Звук: три эффекта из воздуха

Библиотеки звуков не нужны — три рабочих эффекта делаются генератором шума и синусом прямо во время сборки.

ЭФФЕКТ	ИЗ ЧЕГО	ГДЕ СТАВИТЬ
вжух (переход)	розовый шум, срезаны низ и верх, быстрое затухание	на каждом стыке сцен
пинг (появление)	два тона, второй тише, затухание по кривой	когда выезжает карточка
удар (акцент)	низкий тон с резким затуханием	под важной фразой

Громкости, проверенные на слух: переход 0,26 · появление 0,16 · акцент 0,22 · музыка 0,09. Музыка всегда сильно тише голоса — это ошибка новичка номер один.

06 Чеклист перед публикацией

- ☐ Раскадровка просмотрена глазами, а не по превью
- ☐ Лицо резкое во всех сценах, край маски не мылит
- ☐ Личного в кадре нет: чужих имён, сумм, лишних окон
- ☐ Субтитры без слипшихся слов на стыках
- ☐ В каждой сцене сверху идёт процесс, а не статичная картинка
- ☐ Каждый стык прикрыт искрой и переходом

- ☐ Длительность видео совпадает с длительностью звука
- ☐ Обложка читается в маленьком квадрате

07 Длинные ролики: 30 минут и час

Важное, ради чего стоит разбираться: это не про короткие рилсы.

Ровно та же система собирает ролик любой длины — обучающее видео на 30 минут, вебинар на час, разбор с анимацией и оформлением. Меняется не подход, а мелочи:

- ☐ **Режь на части.** Часовое видео собирается кусками по 5–10 минут и склеивается в конце — так проще искать ошибку и не ждать час на каждую правку.
- ☐ **Субтитры на длинном материале** делаются тем же распознаванием: час речи обрабатывается за несколько минут.
- ☐ **Анимация и графика** добавляются теми же карточками — просто их больше.
- ☐ **Проверяй по кускам.** Ошибку в 40-й минуте искать в целом файле мучительно.

Суть не в длине ролика и не в конкретном сервисе. Суть в том, чтобы понимать, кто за что отвечает: кто распознаёт, кто режет, кто озвучивает, кто накладывает. Когда это в голове уложилось, ты соберёшь свою систему под свой формат — хоть рилс, хоть лекцию на час.

08 Если не разобрался

Честно: с первого прочтения это не собирается. У меня первый ролик пересобирался трижды, а вся система выстраивалась неделю. Зато потом каждый следующий ролик готовится сам.

Если застрял — напиши мне в Telegram, оформим консультацию: разберём твой формат, покажу вживую, кто за что отвечает и где обычно ломается. Не хочешь разбираться сам — тоже напиши, соберу систему под тебя.

Связь. Telegram-канал [@alex_aicontent](#) — ссылка на личку в профиле канала. Сайт: alexaistore.ru

Гайд бесплатный. Если пригодился — поделись им с тем, кто до сих пор монтирует руками.