

Что дать AI-агенту для ролика?

Из чего на самом деле собирают ролик:
исследование, исходники, монтаж, графика, звук,
генерация и проверка прав.

Не ищите одну волшебную модель. Сначала назовите задачу: найти
фрагмент, объяснить мысль, починить кадр или собрать голос.

РАБОЧЕЕ МЕСТО АГЕНТА

6 СЛОЁВ

01. Сценарий

research, факты, storyboard, assets

02. Клипнинг

сцены, паузы, хайлайты, титры

03. Графика

HTML, React, Manim, motion-kit

04. Голос и звук

TTS, музыка, SFX, дакинг

05. Кадр и генерация

маски, depth, video, avatars

Сначала задача. Потом инструмент.

Один ролик проходит через несколько разных программ. У каждой - своя работа.

«Сделай видео» - не задача.
«Найди 25 сильных секунд,
выровняй титры и собери 9:16» -
задача.

Метки на карточках показывают: что уже используем, что берём как приём, а что ещё нужно проверить.

[в движении] уже работает в нашем контуре.

[берём приёмы] полезная архитектура или motion-паттерн.

[кандидат] открытая развилка для своей задачи.

[проверка] лицензия кода, весов или зависимостей требует чтения.

Есть длинный исходник

Сцены, тишина, транскрипт,
тайминги, хайлайты.

Нужно объяснить мысль

Программная графика, схема,
текст, ритм.

Кадра нет или он слабый

Маски, апскейл, глубина,
генерация, avatar.

Два вопроса не делегируются. Что считать хорошим роликом и можно ли использовать инструмент в нужном коммерческом сценарии. Первое остаётся редакторским решением. Второе подтверждается первоисточником лицензии.

Минимальный рабочий стек.

Если нужно собрать короткий объясняющий ролик из готового исходника, начните с этих пяти деталей. Видео-генерация здесь не обязательна.

-
- | | | |
|----|---|-------------|
| 01 | WhisperX
Слова, спикеры и точные таймкоды. | РАСШИФРОВКА |
| 02 | auto-editor
Черновой монтаж: тишина и паузы. | ПЕРВЫЙ РЕЗ |
| 03 | FFmpeg
Склейка, кодек, звук, финальный файл. | СБОРКА |
| 04 | HTML / Remotion
Титры, карточки, схемы и объясняющая графика. | ВИЗУАЛ |
| 05 | Qwen3-TTS
Голос, если нет живой записи. | ОЗВУЧКА |

Что остаётся человеку: тезис, факты, отбор фрагментов и финальный просмотр. Именно здесь решается, стоит ли ролик смотреть.

Идея, факты, сценарий и кадры

Сначала тезис и факты. Затем раскадровка, футаж или генерация.

● OpenMontage ↗

ИДЕЯ ДЛЯ СБОРКИ

Исследование до сценария, поиск футажа в открытых архивах, защита от слайдшоу и учёт бюджета. Код AGPL: берём логику процесса, не тащим в закрытый сервис.

○ Story2Board ↗

ПРОВЕРЬ ПРАВА

Раскадровка и locked identity персонажа до генерации. Зависимые модели проверяем отдельно.

○ article_to_scenario ↗

ИСПОЛЬЗУЕМ

Markdown: заголовок становится сценой, список - тезисом, нумерация - схемой.

○ Pollinations ↗

ПРОВЕРЬ ПРАВА

Открытая платформа-агрегатор. Удобна как слой доступа, не как гарантия прав на конкретную модель.

○ TurboReel / MediaChain ↗

СМОТРИМ ПОДХОД

Low-friction генерация через polling и быстрые UGC-сценарии. Берём паттерн, не обещаем стабильный бесплатный провайдер.

○ OpenShorts ↗

МОЖНО ПРОБОВАТЬ

Self-hosted platform: clip generator, AI-shorts и UGC. Полезны PAS-beatsheet и заготовки автосборки.

○ Video Toolkit ↗

ИДЕЯ ДЛЯ СБОРКИ

Cloud GPU, stem separation, text cards, музыка и техническая рутина для агентного видеопроцесса.

Принцип. Выбрать тезис, проверить факт, убрать лишнее - это редакторская работа. Остальное можно разложить на шаги.

01

Факт

первичный источник

→ 02

Тезис

что зритель поймёт

→ 03

Сцена

что показать

→ 04

Кадр

футаж или генерация

Клипинг, понимание, субтитры

Сцены, речь и паузы дают черновой монтаж.
Человек выбирает, что в нём действительно
стоит оставить.

● **video-use** ↗

ИДЕЯ ДЛЯ СБОРКИ

Выбирает сцены, расшифровку и неповторяющиеся кадры вместо прогона всего ролика через модель. Эффекты описываются словами, а команда FFmpeg проходит проверку.

○ **Claude Video** ↗

МОЖНО ПРОБОВАТЬ

Скачать, снять кадры и субтитры, чтобы агент видел референс, а не угадывал по заголовку.

○ **PySceneDetect** ↗

МОЖНО ПРОБОВАТЬ

Режет исходник по сменам сцен.

○ **auto-editor** ↗

МОЖНО ПРОБОВАТЬ

Убирает тишину, сохраняя margin вокруг резов.

○ **X-CLIP** ↗

МОЖНО ПРОБОВАТЬ

Ранжирует футаж по движению и смыслу запроса.

○ **AI YouTube Shorts Gen.** ↗

ПРОВЕРЬ ПРАВА

Виральные хайлайты и content-type detection. Лицензию README надо сверять с файлом LICENSE.

○ **CutClaw** ↗

ТОЛЬКО ИДЕЯ

AMV-монтаж по биту, энергии и сменам музыки. Лицензия не объявлена.

○ **Kadr** ↗

ИДЕЯ ДЛЯ СБОРКИ

GPL-редактор: motion blur и мысль о том, что агент работает рядом с таймлайном.

○ **SupoClip** ↗

AGPL

Self-hosted альтернатива Opus Clip: viral scoring, captions, crop, b-roll. Не зависимость закрытого сервиса.

01

Исходник

длинный ролик

→ 02

Сцены

точки смены

→ 03

Хайлайты

смысл и движение

→ 04

Сборка

титры и 9:16

Слова тоже часть монтажа.

Таймкоды слов нужны для двух вещей: точных титров и графики, которая появляется в нужный момент.

РАСШИФРОВКА И ТИТРЫ

○ Whisper ↗

МОЖНО ПРОБОВАТЬ

Базовая точка качества.

○ faster-whisper ↗

МОЖНО ПРОБОВАТЬ

Быстрая локальная расшифровка.

○ WhisperX ↗

ПРОВЕРЬ ПРАВА

Word-level тайминги и forced alignment.

○ pyannote ↗

ПРОВЕРЬ ПРАВА

Диаризация спикеров.

○ libass / align.py

ИСПОЛЬЗУЕМ

ASS-караоке и точный синк слов.

○ Opus / Submagic / CapCut

ВНЕШНИЕ СЕРВИСЫ

Источник caption-пресетов, b-roll и emoji-паттернов.

Габля из практики. Русские word boundaries у edge-tts могут оказаться пустыми. Синтезатор даёт голос, а истину о времени лучше брать из WhisperX.

01

Файл

речь и паузы

02

Whisper

→ черновой текст

03

WhisperX

→ слова и таймкоды

04

ASS

→ титры в кадре

Графика: HTML, Remotion, Manim

HTML, React и Manim нужны для разных сцен: карточки, интерфейсы, графики и формулы.

● Remotion ↗

ГОТОВЫЙ МОДУЛЬ

React-композиции и данные. Уместен, когда ролик похож на интерфейс, продуктовую демонстрацию или повторяемый шаблон.

○ HTML + Playwright

ИСПОЛЬЗУЕМ

Вертикальная editorial-графика, сетки, SVG, mesh-фоны, CSS-маски. Снимаем headless-браузер покадрово.

○ Pillow

ИСПОЛЬЗУЕМ

Статичная инфографика, карточки и схемы. Дешевле GPU и стабильнее шаблона.

○ Manim ↗

МОЖНО ПРОБОВАТЬ

Графики, формулы, матрицы, камеры, trackers и morphing схем.

○ HyperFrames ↗

ИДЕЯ ДЛЯ СБОРКИ

9 GL-переходов, 16 grades, VarFont, caption DNA, bezier и spring-рецепты.

○ remotion-dev ↗

ИДЕЯ ДЛЯ СБОРКИ

clockWipe, iris, flipIn, rough-notation, rack focus, marquee, starburst и mesh.

○ Video Overlay Kit ↗

МОЖНО ПРОБОВАТЬ

«Центр и спутники», сравнения, lifecycle и strokeDraw.

○ Primer

СМОТРИМ ПОДХОД

Симуляции агентов, SDF/WebGL-диаграммы, blob-ведущий, easeWithPause и mixHsv.

○ lumenshaders ↗

МОЖНО ПРОБОВАТЬ

Ripple и glitch-шейдеры. Эффект работает как пунктуация, не как фон всего ролика.

○ GSAP ↗

НА ВЫБОР

Коммерческое использование сейчас покрыто лицензией. Мы сохраняем нативные APIs для детерминизма.

Практический выбор. HTML - для вертикальной редакционной графики. Remotion - для React-композиций. Manim - для формул, графиков и причинных схем.

Сначала звучит. Потом длится.

Сначала соберите голос. По нему проще выставить длину сцены, титры и паузы.

Qwen3-TTS ↗

МОЖНО ПРОБОВАТЬ

Локальный TTS.

Chatterbox ↗

МОЖНО ПРОБОВАТЬ

Клонирование голоса, watermark и веса сверяем отдельно.

Kokoro ↗

МОЖНО ПРОБОВАТЬ

Компактный TTS.

Piper ↗

GPL

Офлайн-фоллбэк. Лицензии конкретных голосов различаются.

edge-tts / ElevenLabs

СЕРВИС

Сменяемый внешний слой, не open source default.

F5-TTS ↗ / XTTS

NC BECA

Технологически интересны, но не чистый коммерческий default.

МУЗЫКА И SFX

Demucs ↗

МОЖНО ПРОБОВАТЬ

Разделить готовый трек на дорожки.

ACE-Step ↗ / YuE ↗

МОЖНО ПРОБОВАТЬ

Музыка, вокал и SFX.

Magenta RT ↗

МОЖНО ПРОБОВАТЬ

Процедурный звук в реальном времени.

MusicGen / MAGNeT

NC BECA

Код открыт, официальные веса не подходят как коммерческий default.

Stable Audio Open ↗

УСЛОВИЯ

Community License с порогом выручки и атрибуцией.

numpy ambient + ducking

ИСПОЛЬЗУЕМ

Фон и автоматическая просадка музыки под речь.

01

Голос

синтез или запись

02

Длина

→ темп сцены

03

Музыка

→ стемы и SFX

04

Даккинг

→ речь сверху

Слабый кадр или кадра нет.

Если исходника не хватает, сначала попробуйте маску, апскейл или глубину. Генерация - следующий шаг, не первый.

КАДР УЖЕ ЕСТЬ

○ [Real-ESRGAN ↗](#)

МОЖНО ПРОБОВАТЬ

Апскейл. BSD-3-Clause.

○ [RIFE ↗](#)

МОЖНО ПРОБОВАТЬ

Интерполяция кадров.

○ [Depth Anything V2 ↗](#)

BECA

Depth для параллакса. Коммерчески чистая только Small.

○ [SAM 2 ↗](#)

МОЖНО ПРОБОВАТЬ

Маски и трекинг объектов.

○ [BiRefNet ↗](#) / [rembg](#)

BECA

Вырезание фона. Условия зависят от модели.

○ [RestoreFormer++ ↗](#)

APACHE

Восстановление лиц.

НЕ СТАВИМ КАК ЗАКРЫТЫЙ DEFAULT

○ [RVM ↗](#)

GPL

Сильное видеоматирование, но copyleft меняет условия.

○ [GFPGAN ↗](#)

ПРОВЕРИТЬ

Исключения во вложенных зависимостях.

○ [CodeFormer ↗](#)

NC

NTU S-Lab License.

○ [ProPainter ↗](#)

NC

Удаление объектов, но некоммерческие условия.

○ [MatAnyone 2 ↗](#)

NC

Матирование, та же S-Lab лицензия.

○ [Upscayl ↗](#)

AGPL

Хороший desktop app, не автоматическая серверная зависимость.

01

Сохранить

лучший исходник

→ 02

Вытащить

маску или depth

→ 03

Восстановить

качество и движение

→ 04

Генерировать

только недостающий кадр

Кадра нет. Нужен человек в кадре.

Самый быстро меняющийся слой. У каждого проекта отдельно проверяем код, веса и условия API.

● Wan 2.2 ↗

АПАСНЕ

Генерация видео из текста или изображения. Apache-2.0; одна из открытых базовых веток для своей инфраструктуры.

○ Mochi 1 ↗

АПАСНЕ

Ещё один зелёный кандидат.

○ LTX-2 ↗

УСЛОВИЯ

Быстро, но Community License с вырочечным порогом.

○ HunyuanVideo ↗

ТЕРРИТОРИИ

Порог по MAU и территориальные исключения.

○ CogVideoX ↗

РАЗНЫЕ ВЕРСИИ

У 2B и старших веток разные правила.

○ FramePack ↗

ВЕСА

Длинное видео на слабом железе, права на веса не ясны.

○ Higgsfield / MuAPI ↗

API

Один ключ ко многим моделям. Это про скорость запуска, не про open source.

○ OmniHuman на fal ↗

СЕРВЕР

Когда важнее быстрый серверный слой, чем свой GPU-кластер.

○ daVinci-MagiHuman ↗

ВСЁ СТЕК

Кандидат на серверный talking-head. Собственные веса не равны всему pipeline.

01

Код

LICENSE

02

→ Веса

model card

03

→ Условия

terms и пороги

04

→ Человек

consent и disclosure

Лицо - отдельный лицензионный ярус.

Для лица важны три вещи: лицензия, права на веса и согласие человека в кадре.

TALKING-HEAD И LIP-SYNC

○ InfiniteTalk ↗

МОЖНО ПРОБОВАТЬ

Длинные аудиоуправляемые ролики.

○ EchoMimic ↗

МОЖНО ПРОБОВАТЬ

Портрет и звук.

○ SadTalker ↗

БЕСА

Проверяем актуальные веса, не legacy-гайд.

○ MuseTalk ↗

OPENRAIL

Липсинк с use restrictions.

○ LatentSync ↗

OPENRAIL

Коммерция возможна с пропуском ограничений.

○ Hallo2 ↗

ЗАВИСИМОСТИ

High-res ветка тянет CodeFormer.

○ Hunyuan Avatar ↗

УСЛОВИЯ

Условия Tencent наследуются дальше.

○ Wav2Lip ↗

NC

Исторически важный, но не коммерческий default.

○ Sonic

NC

CC-BY-NC-SA, в прод не берём.

Необходимая проверка. Если модель делает лицо реального человека, к лицензии добавляются согласие, раскрытие синтетической природы и правила площадки.

Лицензия - это часть архитектуры.

У одного инструмента могут быть разные правила для кода, весов, зависимостей и результатов.

1. Код

Смотрим LICENSE, а не только бейдж в README. AGPL и GPL меняют условия закрытого сервиса и дистрибутива.

2. Веса

MIT-код не делает веса коммерческими. Типовая ловушка: CC-BY-NC в model card.

3. Зависимости

Проверяем, что докачивается на первом запуске: InsightFace, FLUX, CodeFormer, gated models.

4. Территория и пороги

Выручка группы, MAU, EU/UK exclusions и обязательная атрибуция живут в terms вендора.

Identity stack не складывается в одну строку. [IP-Adapter](#) без FaceID и с FaceID - разные маршруты. [InstantID](#) выпускает research-only checkpoints. [PuLID](#) обычно тянет InsightFace: код MIT, а часть весов non-commercial research only.

Мини-чек перед интеграцией. LICENSE, model card, requirements, terms сайта. Потом отдельно: можно ли запускать, можно ли выпускать продукт, можно ли использовать результат. И обязательно записать дату проверки.

ВНЕШНИЕ СЕРВИСЫ КАК ИСТОЧНИК ПАТТЕРНОВ

Opus Clip, Submagic, CapCut и SupoClip полезны не только как кнопки, а как библиотека UX-приёмов: hormozi/mrbeast captions, emoji, b-roll, virality score, combo-пресеты. Их экран не нужно копировать. Нужно понять, какую монтажную задачу он решает.

Где мы ставим стоп.

Не плохие проекты. Просто не подходят как тихая зависимость в коммерческой сборке.

NC

Wav2Lip

Исторически важный lip-sync, но не коммерческий default.

NC

CodeFormer

Восстановление лиц с некоммерческим ограничением.

NC

ProPainter

Сильный инструмент удаления объектов, но закрывает коммерческий маршрут.

NC

MatAnyone 2

Матирование: та же исследовательская лицензия.

AGPL

Upscayl

Хорошее desktop-приложение, но не серверная зависимость по умолчанию.

RESEARCH

InstantID

Research-only checkpoints: нельзя читать как зелёный свет для продукта.

FACE STACK

IP-Adapter FaceID

Маршрут тянет отдельную проверку InsightFace и его весов.

FACE STACK

PuLID

Та же ловушка: MIT-код не очищает права на лица и веса.

Правило: сначала проверяем LICENSE, model card и зависимости. Потом решаем, появится ли проект в продукте.

Все инструменты остаются кликабельными

Карточки выше ведут в первоисточники. Здесь -
быстрый индекс по слоям.

Фундамент

[FFmpeg](#) · [Pillow](#) · [Playwright](#) · [libass](#) · [MoviePy](#)

Сценарий и исходник

[OpenMontage](#) · [Story2Board](#) · [Pollinations](#) · [OpenShorts](#)

Клипинг и текст

[video-use](#) · [PySceneDetect](#) · [auto-editor](#) · [WhisperX](#) · [faster-whisper](#)

Графика

[Remotion](#) · [Manim](#) · [HyperFrames](#) · [Video Overlay Kit](#) · [lumenshaders](#)

Аудио

[Qwen3-TTS](#) · [Chatterbox](#) · [Kokoro](#) · [Demucs](#) · [ACE-Step](#)

Кадр

[Real-ESRGAN](#) · [RIFE](#) · [Depth Anything V2](#) · [SAM 2](#)

Video и avatars

[Wan 2.2](#) · [LTX-2](#) · [HunyuanVideo](#) · [InfiniteTalk](#) · [MuseTalk](#) · [OmniHuman/fal](#)

Честная граница. Для Primer и MediaChain в исходных заметках не было канонических ссылок. В PDF они отмечены как референсы, а не вымышленные URL.

Полный список ССЫЛОК — 1.

Полные URL, увеличенный кегль. Ссылки можно
открыть в PDF, скопировать или отдать агенту.

↗ https://github.com/m-bain/whisperX	↗ https://github.com/openai/whisper
↗ https://github.com/WyattBlue/auto-editor	↗ https://github.com/SYSTRAN/faster-whisper
↗ https://ffmpeg.org/	↗ https://github.com/pyannote/pyannote-audio
↗ https://www.remotion.dev/	↗ https://github.com/ManimCommunity/manim
↗ https://github.com/QwenLM/Qwen3-TTS	↗ https://github.com/heygen-com/hyperframes
↗ https://github.com/calesthio/OpenMontage	↗ https://github.com/remotion-dev/remotion
↗ https://github.com/DavidDinkevich/Story2Board	↗ https://github.com/alichherawalla/video-overlay-kit
↗ https://github.com/pollinations/pollinations	↗ https://github.com/Leonxlnx/lumenshaders
↗ https://github.com/mutonby/openshorts	↗ https://gsap.com/
↗ https://github.com/digitalsamba/claude-code-video-toolkit	↗ https://github.com/resemble-ai/chatterbox
↗ https://github.com/browser-use/video-use	↗ https://github.com/hexgrad/kokoro
↗ https://github.com/bradautomates/claude-video	↗ https://github.com/OHF-Voice/piper1-gpl
↗ https://github.com/Breakthrough/PySceneDetect	↗ https://github.com/SWivid/F5-TTS
↗ https://github.com/microsoft/VideoX/tree/master/X-CLIP	↗ https://github.com/facebookresearch/demucs
↗ https://github.com/SamurAIGPT/AI-Youtube-Shorts-Generator	↗ https://github.com/ace-step/ACE-Step
↗ https://github.com/GVCLab/CutClaw	↗ https://github.com/magenta/magenta-realtime
↗ https://github.com/HelpFreedom/kadr	↗ https://huggingface.co/stabilityai/stable-audio-open-1.0
↗ https://github.com/FujiwaraChoki/supoclip	

Репозитории, официальная документация и model cards из первой половины карты.

Полный список ссылок — 2.

Продолжение. Полный адрес остаётся текстом и кликабельной ссылкой.

↗ https://github.com/xinntao/Real-ESRGAN	↗ https://muapi.ai/docs/models
↗ https://github.com/hzwer/ECCV2022-RIFE	↗ https://fal.ai/learn/devs/omnihuman-15-user-guide
↗ https://github.com/DepthAnything/Depth-Anything-V2	↗ https://github.com/GAIR-NLP/daVinci-MagiHuman
↗ https://github.com/facebookresearch/sam2	↗ https://github.com/MeiGen-AI/InfiniteTalk
↗ https://github.com/ZhengPeng7/BiRefNet	↗ https://github.com/antgroup/echomimic
↗ https://github.com/wzhouxiff/RestoreFormerPlusPlus	↗ https://github.com/OpenTalker/SadTalker
↗ https://github.com/PeterLln/RobustVideoMatting	↗ https://github.com/TMElyralab/MuseTalk
↗ https://github.com/TencentARC/GFPGAN	↗ https://github.com/bytedance/LatentSync
↗ https://github.com/sczhou/CodeFormer	↗ https://github.com/fudan-generative-vision/hallo2
↗ https://github.com/sczhou/ProPainter	↗ https://github.com/Tencent-Hunyuan/HunyuanVideo-Avatar
↗ https://github.com/pq-yang/MatAnyone2	↗ https://github.com/Rudrabha/Wav2Lip
↗ https://github.com/upscale/upscale	↗ https://github.com/tencent-ailab/IP-Adapter
↗ https://github.com/Wan-Video/Wan2.2	↗ https://github.com/InstantID/InstantID
↗ https://github.com/genmoai/models	↗ https://github.com/ToTheBeginning/PuLID
↗ https://github.com/Lightricks/LTX-2	↗ https://python-pillow.org/
↗ https://github.com/Tencent-Hunyuan/HunyuanVideo	↗ https://playwright.dev/
↗ https://github.com/THUDM/CogVideo	↗ https://github.com/libass/libass
↗ https://github.com/l1llyasviel/FramePack	↗ https://github.com/Zulko/moviepy

Primer и MediaChain не перечислены: в исходных заметках нет канонических URL, и мы не стали их выдумывать.

Сохрани карту. Остальное – в канале.

Under the Hood: разборы AI-индустрии, ссылки на первоисточники и заметки о том, что реально работает в продакшене.



Under the Hood

@gorilla_under_hood ↗



Сканируй QR или открой t.me/gorilla_under_hood