

1 · Три семейства (2026)

Google Gemini, OpenAI GPT-5 и Anthropic Claude Opus 4. Одинаковая интеграция: отправить видео или кадры + вопрос, получить текст, платить за токены. У каждого большой тир «pro» и дешёвый «flash/mini».

Gemini 2.5/3.x : нативное видео | 1M контекст | flash дешево, pro ~\$1.25-2.50 in / \$10-15 out
GPT-5 / 5.4 / 5.5 : нативное видео на 5.4+ | до ~1.1M контекст | ~\$1.25-5 in / \$10-30 out
Claude Opus 4.x : изображения + скриншоты | 1M контекст | \$5 in / \$25 out | уникам: Computer Use

2 · Два счёта, не один

Ввод и вывод счётом отдельно; вывод в разы дороже за токен. Всегда считайте оба.

10-мин видео @ 1 fps, Gemini разрешение по умолчанию:

600 кадров x 258 ток/кадр + 600 с x 32 ток/с = ~174 000 входных токенов

174 000 / 1 000 000 x \$0.30/M (flash) = ~ \$0.052 ввод. Вывод оплачивается сверху.

3 · Контекстное окно — жёсткая стена

Модель на 1 миллион токенов вмещает около 1 часа видео в разрешении по умолчанию или около 3 часов при низком. Сверх этого: понизить разрешение, понизить частоту кадров или перейти к поиску (мультимодальный RAG).

4 · Claude Computer Use — агентный цикл

Цикл из четырёх шагов

1. Вы отправляете скриншот + цель.
2. Claude возвращает действие (нажать/печатать).
3. Ваш код выполняет его в песочнице.
4. Ваш код присылает новый скриншот -> повтор.

Правила безопасности

- Claude не касается машины; это делает ваш код.
- Запуск в песочнице с минимальными правами.
- Без чувствительных данных; allowlist доменов.
- Человек подтверждает значимые действия.

5 · Какое семейство для какой функции

- ☐ Длинное записанное видео в объёме? Gemini flash — зрелое нативное видео, предсказуемая цена за секунду.
- ☐ Видео внутри тяжёлого рассуждения / инструментов? GPT-5 (5.4+) — крупнейшая экосистема, одна сессия.
- ☐ Должна нажимать, печатать, водить экран? Claude Computer Use — единственный продукт автоматизации экрана.
- ☐ Регулируемая, критичная к безопасности сфера? Claude — аккуратные ответы и строгое поведение безопасности.
- ☐ Данные не покидают серверы, или огромный масштаб? Ни одна — open-weights: резидентность, цена, дообучение.
- ☐ Закрепили старую версию? Проверьте: исходный GPT-5 требует извлечения кадров; 5.4+ принимает видео нативно.