

1 · Общий трёхчастный рецепт

Каждая открытая VLM — это «читатель изображений» (визуальный энкодер, напр. CLIP) + маленькая переключатель (проектор/адаптер) + языковая модель. LLaVA доказала в 2023, что обучать надо только маленькую переключатель. Четыре семейства различаются качеством частей и способом обучения, не формой.

2 · Четыре семейства (срез 2026)

Qwen-VL (Qwen3-VL) : Apache 2.0 | 2B-32B dense + 235B MoE | 256K->1M | ЛУЧШАЯ В СРЕДНЕМ
InternVL (InternVL3) : MIT-style | 1B-78B | ~72% MMMU | чистейшая лицензия, топ-качество
Pixtral 12B (Mistral): Apache 2.0 | 12B | 128K контекст | влезает на ОДНУ скромную GPU
LLaVA (OneVision-1.5): Apache 2.0 | полностью открытые данные | эталон, не продакшен

3 · Лицензия обычно важнее бенчмарка

Apache 2.0 (Qwen, Pixtral) и MIT (InternVL) разрешают неограниченное коммерческое использование в любом масштабе. Community-лицензии (у некоторых моделей) разрешают коммерцию, но добавляют условия и потолок гипермасштаба, и являются source-available, а не формально open-source. Модель на два пункта выше, но с лицензией, которую отклонят юристы, не выпустит ничего. Сначала лицензия — потом лидерборд.

4 · Свой хостинг против закрытого API — безубыточность

Арифметика

$\$2/\text{час GPU} \times 24 \times 30 = \$1\,440$ / мес, фикс.
 $\$1\,440 / \$0,05 \text{ за видео} = \sim 28\,800$ видео.
Выше ~29 тыс. видео/мес: своя GPU выигрывает.

Скрытые счета

— Простой GPU: платите полную цену за малую работу.
— Эксплуатация: развернуть, патчить, мониторить.
— vLLM упрощает, но бесплатные веса != бесплатно запускать.

5 · Какая открытая модель под какую задачу

- ☐ Лучшая универсальная открытая модель, безопасный выбор по умолчанию? Qwen3-VL — сильнейшие оценки, Apache 2.0, длинный видеоконтекст.
- ☐ Нужно топ-качество под чистейшей (MIT-style) лицензией? InternVL3.
- ☐ Должно влезть на одну скромную GPU (edge, батч, рядом с камерой)? Pixtral 12B — маленькая и способная.
- ☐ Нужен полностью воспроизводимый исследовательский эталон, не продакшен? LLaVA-OneVision-1.5.
- ☐ Данные могут уходить с серверов, малый объём, нужен быстрый прототип? Арендуйте закрытый API.
- ☐ Для ВИДЕО-продукта взвешивайте видео-бенчмарки (Video-MME), а не изображенческие оценки (MMMU).