

1 - The shared three-part recipe

Every open VLM is an image reader (vision encoder, e.g. CLIP) + a small connector (projector/adaptor) + a language model. LLaVA proved in 2023 that you only have to train the small connector. The four families differ in how good each part is and how they are trained, not in the shape.

2 - The four families (2026 snapshot)

```
Qwen-VL (Qwen3-VL) : Apache 2.0 | 2B-32B dense + 235B MoE | 256K->1M context | BEST ALL-ROUND
InternVL (InternVL3): MIT-style | 1B-78B | ~72% MMMU | cleanest licence, top quality
Pixtral 12B (Mistral): Apache 2.0 | 12B | 128K context | fits on ONE modest GPU
LLaVA (OneVision-1.5): Apache 2.0 | fully open data | research baseline, not production
```

3 - The licence usually beats the benchmark

Apache 2.0 (Qwen, Pixtral) and MIT (InternVL) allow unrestricted commercial use at any scale. Community licences (some other models) allow commercial use but add conditions and a hyperscale cap, and are source-available, not formally open-source. A model two points higher with a licence your lawyers reject ships nothing. Read the licence before the leaderboard.

4 - Self-host vs closed API — the break-even

The arithmetic

\$2/hr GPU x 24 x 30 = \$1,440 / month, fixed.
\$1,440 / \$0.05 per video = ~28,800 videos.
Above ~29k videos/month: own a GPU wins.

The hidden bills

- Idle GPU: paying full price for little work.
- Operations: deploy, patch, monitor, restart.
- vLLM eases serving, but free weights != free to run.

5 - Which open model for which job

- Best all-round open model, the safe default? Qwen3-VL - strongest scores, Apache 2.0, long video context.
- Need top quality under the cleanest possible (MIT-style) licence? InternVL3.
- Must fit on one modest GPU (edge, batch, near a camera)? Pixtral 12B - small and capable.
- Want a fully reproducible research baseline, not production? LLaVA-OneVision-1.5.
- Data can leave servers, low volume, need fast prototype? Rent a closed API instead.
- For a VIDEO product, weight video benchmarks (Video-MME), not still-image scores (MMMU).