Closed-Frontier Model Selection Sheet

Fora Soft Learn · AI for Video Engineering

1 · The three families (2026)

Google Gemini, OpenAI GPT-5, and Anthropic Claude Opus 4. Same integration: send video or frames + a question, get text, pay per token. Each has a big 'pro' tier and a cheap 'flash/mini' tier.

```
Gemini 2.5/3.x : native video | 1M context | flash tier cheap, pro ~$1.25-2.50 in / $10-15 out
GPT-5 / 5.4 / 5.5 : native video on 5.4+ | up to ~1.1M context | ~$1.25-5 in / $10-30 out
Claude Opus 4.x : images + screenshots | 1M context | $5 in / $25 out | unique: Computer Use
```

2 · Two bills, not one

Input and output are billed separately; output is several times pricier per token. Always price both.

```
10-min video @ 1 fps, Gemini default resolution:
  600 frames x 258 tok/frame + 600 s x 32 tok/s = ~174,000 input tokens
  174,000 / 1,000,000 x $0.30/M (flash) = ~ $0.052 input. Output billed on top.
```

3 · The context window is a hard wall

A 1-million-token model holds about 1 hour of video at default resolution, or about 3 hours at low. Beyond that: lower resolution, lower frame rate, or switch to retrieval (multimodal RAG).

4 · Claude Computer Use — the agent loop

The four-step cycle

1. You send a screenshot + a goal.
2. Claude returns an action (click/type/scroll).
3. Your code performs it on a sandbox.
4. Your code returns a new screenshot -> repeat.

Rules of safety

- Claude never touches the machine; your code does.
- Run in a sandboxed VM with minimal privileges.
- Keep credentials out; allowlist domains.
- Human confirms any consequential action.

5 · Which family for which feature

- [ ] Long recorded video at high volume? Gemini flash tier - mature native video, predictable per-second cost.
- [ ] Video inside heavy text reasoning / tool use? GPT-5 (5.4+) - largest tool ecosystem, one unified session.
- [ ] Must click, type, drive a real screen? Claude Computer Use - the only first-class screen-automation product.
- [ ] Regulated, safety-critical vertical? Claude - careful, well-hedged answers and strong safety behaviour.
- [ ] Data cannot leave your servers, or huge scale? None - go open-weights for residency, unit cost, fine-tuning.
- [ ] Pinned an old version? Check it: original GPT-5 needs frame extraction; 5.4+ ingests video natively.