SAM 2 Integration Decision Worksheet

Answer the eight questions before integrating SAM 2 into a video product.

1. Prompt source — where does the click come from?

- ☐ Human in the loop (analyst, doctor, editor)
- ☐ Detector handoff (YOLO / RF-DETR bounding box)
- ☐ Open-vocabulary detector (Grounding DINO / Florence-2)
- ☐ Vision-language agent (Gemini / Claude / GPT)

2. Checkpoint size — quality vs cost

- ☐ Hiera-Tiny (38.9M) — edge / Jetson / cost-bound
- ☐ Hiera-Small (46M) — edge with slightly better quality
- ☐ Hiera-Base+ (80.8M) — cloud sweet spot ($/quality)
- ☐ Hiera-Large (224.4M) — accuracy-bound: VFX / medical

3. Deployment target

- ☐ Cloud (single-stream propagation runs)
- ☐ Cloud (multi-stream, batched)
- ☐ Edge — Jetson Orin Nano / NX / AGX
- ☐ On-device (laptop / desktop with TensorRT FP16)

4. Typical video length and memory-reset cadence

- ☐ Short clips (≤ 300 frames): no reset needed
- ☐ Medium (300–2,000 frames): reset every 300 frames
- ☐ Long (> 2,000 frames): SAM2Long or chunked propagation

5. Multi-object scope

- ☐ Single object per clip — propagation is sufficient
- ☐ A few objects (2–5) — separate propagations per object
- ☐ Many objects (> 5) — consider SAM 3 (concept segmentation)

6. Occlusion strategy

- ☐ Threshold occlusion score (suggested: 0.7)
- ☐ Emit empty mask when above threshold
- ☐ Surface low-confidence frames for human review

7. Licence audit

- ☐ Apache 2.0 on model weights — commercial OK
- ☐ CC BY-SA on SA-V data — copyleft on fine-tunes
- ☐ If fine-tuning, document data lineage at publish time

8. Fallback when SAM 2 fails

- ☐ Distractor lock-on: switch to SAM2.1++ memory
- ☐ Long-video drift: SAM2Long or chunked re-prompt
- ☐ Edge fidelity (hair / thin objects): hand-off to manual
- ☐ Multi-instance concept: escalate to SAM 3