# VLM Fine-Tuning Readiness Checklist

Fora Soft Learn · AI for Video Engineering

## 1 · Do you even need to fine-tune? (try in order)

- [ ] Prompt first: a clear instruction + 3 good examples in the request. Free, an afternoon. If it works, STOP.
- [ ] Retrieval (RAG) next: the gap is missing facts that change often (rosters, policies, recent events). Update in seconds, no training.
- [ ] Fine-tune only for BEHAVIOUR: a specialised vocabulary, a fixed output format, a domain visual judgment, or very high call volume.
- [ ] Strongest systems do all three. Retrieval + fine-tuning together beat either alone (>11 points in 2026 tests).

## 2 · Method + GPU-memory cheat-sheet (7B model)

| | | |
|---|---|---|
| **Full fine-tune (16-bit)** | ~120 GB | most flexible, most expensive, HIGH forgetting risk |
| **Full fine-tune (8-bit)** | ~60 GB | still multi-GPU territory |
| **LoRA (16-bit)** | ~16 GB | THE DEFAULT — frozen original + small overlay, low forgetting |
| **QLoRA (4-bit)** | ~6 GB | cheapest; cannot train the vision tower at 4-bit |

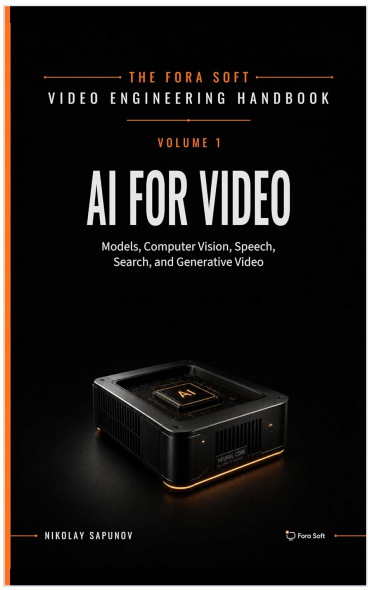
## 3 · Dataset — the real cost, not the training run

- [ ] A few hundred examples teach a format or modest vocabulary; a new visual judgment needs the low thousands.
- [ ] Quality beats quantity: 1,000 clean, varied, correctly-labelled clips beat 10,000 sloppy ones.
- [ ] Watch for shortcuts: if every 'intruder' clip is at night, the model learns 'night', not 'intruder'.
- [ ] Budget the humans: ~1,000 clips at ~20/hr labelling + review + a test set = ~75 person-hours. The GPU run is hours and $10-30.

## 4 · Measure before you ship (the forgetting check)

- [ ] Split the data: hold back 10-20% as a test set the model NEVER sees during training.
- [ ] Score the ORIGINAL model on a few general tasks BEFORE fine-tuning; write the numbers down.
- [ ] After fine-tuning, re-run those general tasks. Domain up + general collapsed = you over-trained.
- [ ] Fix over-training: fewer epochs, use LoRA not full fine-tune, or mix some general examples back in.

Companion to: 'Fine-Tuning A Video VLM On A Domain'
www.forasoft.com/learn/ai-for-video-engineering/articles-ai/finetune-vlm-on-domain-surveillance-telemedicine




GO DEEPER · VOLUME 6 OF 7

# AI for Video: A Complete Guide to Intelligence in Video Products

From Models and Inference to Computer Vision, Speech, Search, and Generative Video

By Nikolay Sapunov, CEO at Fora Soft · 1,444 pages · Kindle & paperback on Amazon

**Read it on Amazon →** amazon.com/dp/B0HH6TKPZ1

This checklist tries prompting and retrieval first, then sizes LoRA, the dataset and the forgetting check for a domain VLM. Volume 6 covers the model side in depth: the ML minimum for product decisions, model artifacts, vision-language models, and RAG over a video archive.

## WHAT YOU WILL FIND INSIDE

- ML basics for product decisions
- Vision-language models for video
- Model artifacts and fine-tuned weights
- RAG over a video archive

**THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON**

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI
Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF
Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery
Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing
Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs
**Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video**
Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

## Thinking about fine-tuning a VLM on your own footage?

We will test prompting and retrieval first, build a labelled set if tuning is needed, and run the forgetting check before release. 250+ video products shipped since 2005.

**Book a 30-min call →** **WhatsApp →** **Email us →**