

# Anomaly Detection Algorithm Selection Worksheet

Run twelve algorithms through the four-number sieve. Land on two or three before the meeting ends.

## 1. The four numbers — fill these in first

<b>Labelled examples</b> 0 → AE / OCSVM / iForest. 2-100 → CLIP-based. 100-1k → MIL. 10k+ → fully supervised. <hr/> your answer	<b>Latency budget</b> <200 ms → edge. 200-500 ms → hybrid. >500 ms → cloud. >10 s → forensic VLM. <hr/> your answer	<b>Per-camera hardware</b> <\$250 → INT8 only. \$250-1k → Orin Nano/NX. >\$1k → cloud GPU. CPU sidecar → watchdog. <hr/> your answer	<b>Explanation need</b> None → skip Layer 4. Review queue → small VLM. Regulator → frontier VLM. Court-grade → Holmes-VAD. <hr/> your answer
------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------

## 2. The twelve algorithms — pick by the sieve, not the marketing copy

#	Algorithm	Labels	Latency	Layer	Watch for
1	One-Class SVM	0	<10 ms	5 / L1	Concept drift — retrain quarterly
2	Isolation Forest	0	<1 ms	5	Slow drift uncaught — anchor baseline
3	Local Outlier Factor	0	10-50 ms	5 (rare)	O(n <sup>2</sup> ) training — prefer iForest
4	Vanilla autoencoder	0	15-25 ms	1	Threshold drift — rolling percentile
5	ST-AE / MemAE	0	80-150 ms	1	Overfit-to-recent — multi-month train
6	VAE / GAN	0	30-100 ms	1 (rare)	Training instability — usually skip
7	3D-CNN (I3D, SlowFast)	100+	25-40 ms	2	Clip-boundary effects — use overlap
8	Video transformer (VideoMAE, Swin)	1k+	60-200 ms	2 (cloud)	HP sensitivity — start pretrained
9	Weakly-supervised MIL (RTFM/MGFN/MTFL)	video-lvl	5-15 ms	2	Bag-level confusion — balance bags
10	CLIP-based (VadCLIP, TPWNG, TrCLIP)	0-100	50-500 ms	3	Prompt brittleness — ensemble prompts
11	Multimodal LLM (Holmes-VAD, Cerberus)	0	200 ms-10 s	4	Hallucination — structured outputs
12	LSTM on engineered features	0	<1 ms	5	Feature engineering is the work

## 3. Wire the shortlist into the five-layer stack

- |                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Layer 1 candidate filter — vanilla AE or MemAE on Orin Nano / Hailo-8.<br><input type="checkbox"/> Layer 2 classifier — SlowFast + RTFM/MGFN/MTFL MIL head, INT8 at the edge.<br><input type="checkbox"/> Layer 3 re-scorer — VadCLIP or TPWNG in the cloud, called on borderline 5-20%. | <input type="checkbox"/> Layer 4 explanation — Qwen-VL 7B for routine + frontier VLM on top 1%.<br><input type="checkbox"/> Layer 5 watchdog — Isolation Forest + LSTM on 60 engineered features.<br><input type="checkbox"/> Integration — ONVIF Profile M out · MQTT bus · audit log to immutable storage. |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## 4. The four traps that kill the budget

<p><b>Starting at the heaviest algorithm</b></p> <p>"Let's use GPT-5 for anomaly detection." Cost runs to \$300/cam/month.                  Fix: start at Layer 1 (cheapest); escalate only when forced.</p>	<p><b>Skipping the watchdog</b></p> <p>Vision layers fire "all clear" while the camera is frozen on a still image.                  Fix: Isolation Forest + LSTM on pipeline features as Layer 5 sidecar.</p>
<p><b>Treating anomaly detection as object detection</b></p> <p>Bolting "alert if person in zone" on YOLO. Misses every temporal class.                  Fix: start from Layer 2 (3D-CNN/MIL); add YOLO for localisation only.</p>	<p><b>Hand-crafted CLIP prompts</b></p> <p>Single hand-written prompt is brittle; AUC swings 3-5 points on rewording.                  Fix: ensemble prompts + learned prompt tuning (CoOp, MaPLe).</p>