

Open-Vocabulary Detection Decision Worksheet

Answer the eight questions before picking between Grounding DINO, Florence-2, OWLv2, RT-DETR, and RF-DETR.

1. Latency budget per frame

- ☐ Under 10 ms — closed-vocab only (RF-DETR-Large, RT-DETRv3).
- ☐ 10 to 50 ms — closed-vocab on hot path, OVD async.
- ☐ 50 to 200 ms — OWLv2 / Florence-2 viable on motion gate.
- ☐ Over 200 ms — Grounding DINO 1.5 Pro / 1.6 Pro acceptable.

2. Accuracy floor (COCO-equivalent AP)

- ☐ Highest possible (60+ AP) — RF-DETR-2XL (closed-vocab).
- ☐ State-of-the-art zero-shot (55+ AP) — Grounding DINO 1.6 Pro.
- ☐ Good-enough zero-shot (30 to 40 AP) — OWLv2 or Florence-2.
- ☐ Fine-tune target — Florence-2 (multi-task) or OWLv2.

3. Prompting interface — what the operator types

- ☐ Free-form description / referring expression — Grounding DINO.
- ☐ List of class-name phrases — OWLv2 or YOLO-World.
- ☐ Task token plus free text — Florence-2 (`` , ``).
- ☐ Fixed class list at training time — RT-DETR / RF-DETR.

4. Licence path

- ☐ Apache 2.0 closed-source — RF-DETR core / OWLv2 / RT-DETR.
- ☐ MIT closed-source — Florence-2 (preferred for commercial).
- ☐ Grounding DINO 1.5 / 1.6 Pro — IDEA Research API contract.
- ☐ Pre-launch legal sign-off complete (week 1, not week 12).

5. Deployment hardware

- ☐ Cloud NVIDIA GPU (T4, L4, A10, A100, H100) — any model.
- ☐ Jetson AGX Orin edge — RF-DETR / RT-DETR or Grounding DINO Edge.
- ☐ CPU-only edge — RT-DETRv3-R18 or RF-DETR-Nano.
- ☐ Apple Silicon — Florence-2 (CoreML export), no FlashAttention.
- ☐ Browser (WebGPU) — YOLO-World or RF-DETR-Nano via WASM.

6. Multi-task scope — beyond detection

- ☐ Detection only — RF-DETR or RT-DETR.
- ☐ Detection + segmentation — RF-DETR (built in) or SAM 2 stack.
- ☐ Detection + OCR + captioning + grounding — Florence-2.
- ☐ Detection + visual grounding referring expressions — Grounding DINO.

7. Auto-labelling plan

- ☐ Two-tier pipeline: closed-vocab hot path + OVD fallback.
- ☐ Prompt-template layer rewrites operator input before model.
- ☐ Confidence threshold tuned per prompt template.
- ☐ Auto-labelled examples reviewed before retraining (>=500 / class).

8. Retraining cadence — closed-vocab refresh

- ☐ Monthly retrain on auto-labelled data (default).
- ☐ Validation set frozen — no leakage from auto-labels.
- ☐ A/B test: new closed-vocab model in shadow for 2 weeks.
- ☐ Rollback path defined if accuracy drops on legacy classes.

The eight-number summary

Latency budget · accuracy floor · prompt interface · licence path · deployment hardware · multi-task scope · auto-labelling plan · retraining cadence.

If any of these eight is unanswered, the open-vocabulary detection decision will be made for the wrong reason. Answer all eight before shipping the first prompt.