```yaml
system:
  name: TSAR_RAPTOR_V6
  role: non_blocking_detection_subsystem
  philosophy:
    - decision_over_label
    - uncertainty_is_asset
    - attention_is_governed
    - fail_open_not_fail_close

hardware_envelope:
  cpu:
    max_parallel_jobs: 32
    soft_cap: 24
  gpu:
    max_concurrent_batches: 4
    soft_cap: 2
    vram_soft_limit: 0.85

safety_valve:
  modes:
    - NORMAL
    - DEGRADED_LOW_ONLY
    - BYPASS
  manual_override:
    enabled: true
    default_mode: NORMAL
    ttl_minutes: 180
    audit_log: true

  auto_circuit_breaker:
    triggers:
      gpu_pressure:
        condition: "gpu.vram_utilization > 0.92"
        action: DEGRADED_LOW_ONLY
      latency_violation:
        condition: "p95_latency_ms > slo.p95_latency_ms"
        action: BYPASS
      human_budget_exhausted:
        condition: "human.today_escalations >= human.max_daily_samples"
        action: DEGRADED_LOW_ONLY
      low_information_period:
        condition: "rolling_1h.mean_information_gain <
policy.minimum_viable_ig"
        action: DEGRADED_LOW_ONLY

slo:
  p95_latency_ms: 300
  timeout_rate: 0.005

  gpu_budget_share: 0.20

  cpu_budget_share: 0.25

  max_high_cost_rate: 0.05


sla:

  non_blocking: true

  fail_mode: FAIL_OPEN

  enforce_resource_caps: true


scheduler:

  states:


    INIT:

      on_enter:

        - reset_cost

        - reset_uncertainty

      next: LOW_COST


    LOW_COST:

      modules: low_cost

      evaluate:

        uncertainty_remaining: U0_remain

        uncertainty_reduction: U0_reduce
        information_gain: I0
      decision:
        - if: "I0 < policy.min_continue"
          next: STOP
        - else: MID_COST

    MID_COST:
      modules: mid_cost
      evaluate:
        uncertainty_remaining: U1_remain
        uncertainty_reduction: U1_reduce
        disagreement: D1
        information_gain: I1
      decision:
        - if: "I1 < policy.min_continue"
          next: STOP
        - if: "U1_remain < policy.low_exit and D1 < policy.disagreement_high"
          next: ROUTE
        - else: HIGH_COST

    HIGH_COST:
      modules: high_cost
      evaluate:
        uncertainty_remaining: U2_remain
        uncertainty_reduction: U2_reduce
        disagreement: D2
      decision:
        - if: "D2 >= policy.disagreement_high"
          next: HUMAN_ESCALATION
        - else: ROUTE

    HUMAN_ESCALATION:
      constraints:
        max_daily_samples: 200
      output:
        - hypothesis_adjustment
        - module_trust_feedback
      terminal: true

    ROUTE:
      description: decision_only_non_label
      output:
        - routing_decision
        - confidence_estimate
        - uncertainty_remaining
        - disagreement_score
      terminal: true
```

```
    STOP:
      output:
        - stop_reason
        - info_density_estimate
        - cost_spent
      terminal: true

policy:
  min_continue: 0.10
  low_exit: 0.15
  disagreement_high: 0.35

policy_learning:
  enabled: true
  objective:
    maximize:
      - uncertainty_reduction_per_cost
    minimize:
      - wasted_high_cost_calls
  adjustable_thresholds:
    min_continue: [0.05, 0.25]
    disagreement_high: [0.25, 0.50]
  safety:
    max_delta_per_update: 0.05
    rollback_on_regression: true

attention_budget_governance:
  core_integrity:
    floor_gpu: 1
    priority: 10
  detection_layer:
    ceiling_gpu: 1
    throttleable: true
    priority: 7
  economic_layer:
    ceiling_gpu: 2
    priority: 5
  degradation_order:
    - economic_layer
    - detection_layer
    - core_integrity
```