

Operating Context Canvas

The Operating Context Canvas is the first artifact produced for any agent build. It locks the scope, risk class, and irreversible-impact boundary before any prompt, tool, or model decision is made. Every downstream S1 artifact (failure modes, thresholds, fallbacks, HITL policy) reads from this canvas. If a field is unknown, mark it UNKNOWN and treat the agent as risk_class: high until resolved.

Fillable fields

```
agent_id: <kebab-case unique id>
agent_objective: <one sentence; the single decision or output the agent must produce>
primary_user: <role; not a person>
risk_class: <low | medium | high | critical>
operating_jurisdiction: <state/locality/regulator if any>
hard_constraints:
  - <statute, contract, or policy the agent must not violate>
  - <data residency, retention, or disclosure rule>
  - <budget or rate-limit ceiling>
expected_inputs:
  - source: <connector/system of record>
    schema: <pointer to schema or example>
    confidence_band_default: <high | medium | low | unknown>
acceptable_outputs:
  - type: <draft | recommendation | filed-action | notification>
    consumer: <who or what receives it>
    reversibility: <reversible | partially-reversible | irreversible>
irreversible_impact_boundary:
  - <action the agent must never take autonomously>
  - <action that always requires an HITL gate>
out_of_scope:
  - <adjacent task the agent must refuse>
escalation_targets:
  - role: <human reviewer role>
    channel: <how they are reached>
    sla_minutes: <response expectation>
success_signals:
  - <observable metric proving the agent worked>
known_unknowns:
  - <data gap, vendor dependency, or legal ambiguity>
sunset_criteria: <what would cause this agent to be retired>
```

Worked example: Stormwater Permit Triage Agent

A real intake agent for a Florida special district. The agent reads permit applications submitted via the district portal, classifies them by stormwater impact, and either routes to a clerk queue or flags for engineer review.

```
agent_id: stormwater-permit-triage
agent_objective: Classify each incoming stormwater permit application into {clerk-queue,
engineer-review, return-to-applicant} and produce a one-paragraph rationale.
primary_user: District clerk
risk_class: high
operating_jurisdiction: South Walton County, FL; FDEP delegated authority
hard_constraints:
  - Must cite the specific FS chapter or DEP rule when recommending return-to-applicant
  - Cannot issue, deny, or sign any permit
  - Cannot transmit applicant PII to any non-district connector
  - Must complete classification within 4 business hours of intake
```

```
expected_inputs:
  - source: district-portal-submissions
    schema: permit-application-v3.json
    confidence_band_default: medium
  - source: parcel-gis-layer
    schema: parcel-feature-v1
    confidence_band_default: high
acceptable_outputs:
  - type: routing-recommendation
    consumer: clerk-queue or engineer-review-queue
    reversibility: reversible
  - type: applicant-notification-draft
    consumer: clerk (for review before send)
    reversibility: reversible
irreversible_impact_boundary:
  - Sending any communication directly to an applicant without clerk approval
  - Modifying parcel or permit records of record
  - Posting to the public agenda or meeting packet
escalation_targets:
  - role: District Engineer
    channel: ops-stormwater Slack + email
    sla_minutes: 240
  - role: Commissioner-on-call
    channel: SMS
    sla_minutes: 60
success_signals:
  - Engineer agreement rate on routing >= 92% over rolling 30 days
  - Median triage latency < 2 business hours
  - Zero direct-to-applicant sends without clerk approval
known_unknowns:
  - Wetlands overlay data refresh cadence from FWC
  - Whether ARPA stormwater grant changes intake fields mid-cycle
sunset_criteria: New state permit portal mandates direct integration replacing district
intake.
```

Usage notes

Fill the canvas before the prompt. The irreversible_impact_boundary list is the seed set for the HITL policy gates and the adversarial test plan. The expected_inputs.confidence_band_default values feed the threshold spec. Treat the canvas as living — re-review on every risk-class change, connector change, or jurisdictional ruling. Version it; do not edit in place after the first production deploy.