

## S1 — Define primary use case

### Epic / Feature

Business Alignment

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

None (kickoff)

### Business Value

Align scope and outcomes; measurable ROI

### Persona

product owner

### Assumptions / Risks

Metrics agreed with stakeholders; Scope creep /  
vague measures

**Story** *As a product owner, I want a single high-value LLM use case defined with success metrics.*

### Non-Functional

Performance

Security

Reliability

Privacy

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** stakeholders and metrics framework are available

**When** we define one use case, objectives, and guardrails

**Then** the use case, metrics, and scope are documented and approved

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Document target personas and jobs-to-be-done
- ☐ Draft success metrics (accuracy/CSAT/cost)
- ☐ Get approval from product/security/finance

## S2 — Risk & compliance profile

### Epic / Feature

Business Alignment

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S1

### Business Value

Clear compliance posture before build

### Persona

security lead

### Assumptions / Risks

Use standard DPA/PIA templates; Unmitigated model/data risks

**Story** *As a security lead, I want a risk register for LLM usage so controls and reviews are clear.*

### Non-Functional

Security

Privacy

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** use case and data classes are known

**When** we run DPA/PIA and classify data & model usage

**Then** the risk register and control owners are recorded

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.

**Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Complete DPA/PIA
- ☐ Define data classification & retention
- ☐ Record model usage constraints

### S3 — Value hypothesis & ROI guardrails

**Epic / Feature**

Business Alignment

**Priority / Estimate**

Priority: Must    SP: TBD

**Dependencies**

S1

**Business Value**

Prevent runaway spend; ensure ROI

**Persona**

FinOps lead

**Assumptions / Risks**

Token pricing and expected volumes estimated;

Budget blowout if underestimated

**Story** *As a FinOps lead, I want a cost envelope and ROI hypothesis so we can enforce budget caps.*

**Non-Functional**

Performance

Cost

**Acceptance Criteria (BDD)**

**Scenario** Happy path

**Given** pricing and volume assumptions are available

**When** we model cost/req, monthly caps, and alert thresholds

**Then** budget envelopes and alerts are configured

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

#### Tasks

- ☐ Set per-request target cost
- ☐ Define monthly cap and alerts
- ☐ Publish ROI hypothesis doc

## S4 — Source inventory

### Epic / Feature

Knowledge Base

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S1

### Business Value

Grounded answers; fewer hallucinations

### Persona

knowledge engineer

### Assumptions / Risks

Identify owners and refresh cadences; Orphaned or stale sources

**Story** *As a knowledge engineer, I want a catalog of authoritative data sources so retrieval is grounded.*

### Non-Functional

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** major content repositories are accessible

**When** we list sources with access, owners, and freshness

**Then** the catalog is published and reviewed

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ List all corp sources
- ☐ Capture owners/access methods
- ☐ Record refresh cadence

## S5 — PII/Secrets policy for corp content

### Epic / Feature

Knowledge Base

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S4

### Business Value

Prevent sensitive-data leakage

### Persona

security lead

### Assumptions / Risks

Redaction/deny-list patterns in place;

Over/under-redaction tradeoffs

**Story** *As a security lead, I want PII/secrets handling rules so ingestion is safe.*

### Non-Functional

Security

Privacy

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** source types and sensitivity are known

**When** we define redaction/deny-list and hashing policy

**Then** policy is approved by security and applied in pipelines

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Write redaction/deny-list patterns
- ☐ Define hashing/retention rules
- ☐ Add policy checks to ETL

## S6 — Ingestion pipeline v1

**Epic / Feature**

Knowledge Base

**Priority / Estimate**

Priority: Must    SP: TBD

**Dependencies**

S4, S5

**Business Value**

Reliable, auditable ingestion

**Persona**

platform engineer

**Assumptions / Risks**

Use idempotent, incremental loads; Silent failures / duplicates

**Story** *As a platform engineer, I want a repeatable ETL to a doc store so content is normalized.*

**Non-Functional**

Reliability

Performance

**Acceptance Criteria (BDD)**

**Scenario** Happy path

**Given** target doc store and credentials are available

**When** we implement incremental loads with checksums and alerts

**Then** new docs appear and failures alert on-call

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Implement incremental ETL
- ☐ Add checksum/versioning
- ☐ Add failure alerts to on-call

## S7 — Chunking & metadata strategy

### Epic / Feature

Knowledge Base

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S6

### Business Value

Better retrieval quality

### Persona

retrieval engineer

### Assumptions / Risks

Baseline chunk size & tags defined; Too-small raises cost / too-large hurts recall

**Story** *As a retrieval engineer, I want chunk sizes and metadata tagging so recall/precision are tunable.*

**Non-Functional** Performance Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** normalized content exists

**When** we apply chunking and tags (titles, timestamps, permissions)

**Then** retrieval quality baselines are recorded

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Configure chunker
- ☐ Tag docs with titles/timestamps/permissions
- ☐ Run baseline retrieval evaluation

## S8 — Model access baseline

### Epic / Feature

LLM Platform

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S1

### Business Value

Unblock experimentation

### Persona

platform engineer

### Assumptions / Risks

Keys and regions set; Rate limits / region constraints

**Story** *As a platform engineer, I want access to chosen base/model(s) so experiments can run.*

**Non-Functional** Performance Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** provider account and roles exist

**When** we provision keys/roles and record rate limits

**Then** experiments can call the model successfully

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/allly checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Create API keys/roles
- ☐ Document rate limits/regions
- ☐ Smoke-test completions



## S9 — Embedding store & similarity search

### Epic / Feature

LLM Platform

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S6, S7

### Business Value

Enable retrieval-augmented generation

### Persona

retrieval engineer

### Assumptions / Risks

Choose embedding model & dims; Hot partitions  
/ index drift

**Story** *As a retrieval engineer, I want an embeddings index so we can RAG.*

### Non-Functional

Performance

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** embedding model and store are selected

**When** we build the index with upsert/delete APIs

**Then** top-k search returns expected neighbors

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Create embeddings index
- ☐ Expose upsert/delete APIs
- ☐ Test top-k similarity queries

## S10 — Secrets & config management

### Epic / Feature

LLM Platform

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S6, S8

### Business Value

Hardened operations

### Persona

devops engineer

### Assumptions / Risks

Use KMS/secret store, least-privilege roles; Secret sprawl / misconfig

**Story** *As a devops engineer, I want secrets & configs centralized so pipelines are safe.*

### Non-Functional

Security

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** KMS/secret store is available

**When** we centralize secrets/configs and review roles

**Then** secrets are rotated and access is audited

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Wire KMS/secret store
- ☐ Refactor services to consume secrets
- ☐ Review IAM roles/permissions

## S11 — Input/output filtering

**Epic / Feature**

Trust & Safety

**Priority / Estimate**

Priority: Must SP: TBD

**Dependencies**

S8, S10

**Business Value**

Reduce abuse and leakage

**Persona**

security lead

**Assumptions / Risks**

Deny/allow lists + ML filters; Overblocking harms UX

**Story** *As a security lead, I want filters for PII, profanity, and malware prompts so we reduce abuse.*

**Non-Functional** Security Privacy

**Acceptance Criteria (BDD)**

**Scenario** Happy path

**Given** routing path for I/O is instrumented

**When** we enforce filters with tests

**Then** unsafe content is blocked and logs show actions

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Implement input filters
- ☐ Implement output filters
- ☐ Add unit tests & sampling

## S12 — Prompt-injection defenses

### Epic / Feature

Trust & Safety

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S11

### Business Value

Safer prompt handling

### Persona

app engineer

### Assumptions / Risks

Bind tools; separate retrieved context; Over-isolation reduces utility

**Story** *As an app engineer, I want system prompts & retrieval isolation so jailbreaks are minimized.*

### Non-Functional

Security

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** tool scopes and retrieval path are known

**When** we enforce content provenance and tool scopes

**Then** injections in the golden set fail to jailbreak

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Add system prompt patterns
- ☐ Isolate retrieval context
- ☐ Test jailbreak suite

## S13 — Data leakage prevention

### Epic / Feature

Trust & Safety

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S11

### Business Value

Meet contractual/regulatory obligations

### Persona

compliance officer

### Assumptions / Risks

Set “no-training” flags with vendors; Incorrect vendor settings

**Story** *As a compliance officer, I want policies preventing training on sensitive user data so we meet obligations.*

### Non-Functional

Security

Privacy

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** vendor controls support no-training

**When** we set flags and capture audit proof

**Then** auditable evidence is available

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Set vendor no-training flags
- ☐ Verify logs/audit proof
- ☐ Document policy in runbooks

## S14 — Golden datasets

### Epic / Feature

Quality & Testing

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S1, S11

### Business Value

Measurable quality

### Persona

QA lead

### Assumptions / Risks

Peer-review labels; cover hard cases; Label noise  
/ poor coverage

**Story** *As a QA lead, I want labeled Q&A/tasks covering happy/edge/jailbreak cases so we can score quality.*

### Non-Functional

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a labeling rubric is ready

**When** we create versioned gold sets

**Then** a coverage report is produced

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Write coverage rubric
- ☐ Label happy/edge/jailbreak cases
- ☐ Version and store dataset

## S15 — Metrics & scoring

### Epic / Feature

Quality & Testing

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S14

### Business Value

Track quality and cost regressions

### Persona

data scientist

### Assumptions / Risks

Automated scoring jobs; Metric mismatch with UX

**Story** *As a data scientist, I want metrics (groundedness, factuality, toxicity, latency, cost) so we track regressions.*

### Non-Functional

Performance

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** gold sets exist

**When** we implement batch scoring with thresholds

**Then** metrics dashboard shows runs and pass/fail status

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Implement scorers
- ☐ Define thresholds for pass/fail
- ☐ Publish evaluation dashboard

## S16 — CI eval gate

### Epic / Feature

Quality & Testing

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S15

### Business Value

Prevent regressions from shipping

### Persona

devops engineer

### Assumptions / Risks

Hook CI to eval harness; Flaky gates break velocity

**Story** *As a devops engineer, I want pre-merge eval jobs so bad prompts/models don't ship.*

### Non-Functional

Reliability

Performance

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** CI system supports checks

**When** we add a pre-merge evaluation check

**Then** PR fails when metrics regress beyond thresholds

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.

**Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Add PR check to CI
- ☐ Fail build on threshold drop
- ☐ Store reports as artifacts



## S17 — RAG query pipeline

### Epic / Feature

Retrieval-Augmented App

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S9, S12

### Business Value

Grounded, cited answers

### Persona

app engineer

### Assumptions / Risks

Top-k with citations; fallback on no context;  
Latency/no-answer cases

**Story** *As an app engineer, I want retrieve→rank→generate flow so answers are cited and grounded.*

### Non-Functional

Performance

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** retrieval and model access are ready

**When** we implement retrieve→rank→generate with citations

**Then** answers cite sources and return empty when low confidence

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.

**Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Implement retrieve→rank→generate
- ☐ Return citations
- ☐ Add no-context fallback

## S18 — Prompt templates & system prompts

### Epic / Feature

Retrieval-Augmented App

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S11, S16

### Business Value

Consistent tone and safer outputs

### Persona

prompt engineer

### Assumptions / Risks

Version prompts with rollback; Template rot / drift

**Story** *As a prompt engineer, I want reusable templates so style and safety are consistent.*

### Non-Functional

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a templating path exists

**When** we version templates and add A/B hooks

**Then** rollback path is validated

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Create template registry
- ☐ Add A/B hooks
- ☐ Test rollback path

## S19 — Re-ranking & hybrid search

### Epic / Feature

Retrieval-Augmented App

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S17

### Business Value

Higher-quality retrieval

### Persona

retrieval engineer

### Assumptions / Risks

Use NDCG/Recall on gold set; Overfitting to gold set

**Story** *As a retrieval engineer, I want BM25+vector hybrid and re-rankers so relevance improves.*

### Non-Functional

Performance

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a hybrid index is in place

**When** we add a re-ranker and measure offline gains

**Then** NDCG/Recall improve vs baseline

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Enable BM25+vector hybrid
- ☐ Add re-ranker
- ☐ Report NDCG/Recall gains

## S20 — Tool adapters

### Epic / Feature

Agent/Tool Use

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S17

### Business Value

Enable safe tool use

### Persona

app engineer

### Assumptions / Risks

Typed schemas, timeouts, auth; Tool misuse /  
privilege escalation

**Story** *As an app engineer, I want safe wrappers for tools (search, calculators, internal APIs) so the LLM can act.*

### Non-Functional

Security

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a tool registry is available

**When** we implement adapters with validation and timeouts

**Then** tool calls succeed or fail safely with logs

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.

**Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Define adapter interfaces
- ☐ Validate/timeout tool calls
- ☐ Enforce per-tool authorization

## S21 — Function calling policy

### Epic / Feature

Agent/Tool Use

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S20

### Business Value

Controlled capability; prevent scope creep

### Persona

security lead

### Assumptions / Risks

Allowlist, rate limits, arg validation; Shadow APIs / over-broad access

**Story** *As a security lead, I want function-calling constraints so scope creep is prevented.*

### Non-Functional

Security

Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a tool inventory is known

**When** we enforce allowlists and quotas

**Then** abuse attempts are blocked and logged

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Set function allowlist
- ☐ Add rate limits
- ☐ Log and alert on violations

## S22 — Structured logging

### Epic / Feature

Telemetry

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S17, S20

### Business Value

Faster debugging; cost control

### Persona

SRE

### Assumptions / Risks

PII-safe structured logs; Sensitive fields leakage risk

**Story** *As an SRE, I want request/response/latency/cost logs so issues are diagnosable.*

### Non-Functional

Reliability

Privacy

Cost

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a log sink is available

**When** we add structured logs with correlation IDs

**Then** dashboards show latency/cost/error trends

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Add correlation IDs
- ☐ Scrub PII fields
- ☐ Publish latency/cost dashboards

## S23 — Tracing

### Epic / Feature

Telemetry

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S22

### Business Value

Identify bottlenecks end-to-end

### Persona

SRE

### Assumptions / Risks

Use APM tracing and span attributes; Missing propagation

**Story** *As an SRE, I want spans for retrieve→rank→generate→tools so bottlenecks are visible.*

### Non-Functional

Reliability

Performance

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** an APM is available

**When** we add spans for each stage

**Then** slow steps are identifiable in traces

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Instrument retrieve/rank/generate/tools
- ☐ Propagate trace context
- ☐ Create trace dashboards

## S24 — Cost dashboards & budgets

### Epic / Feature

Telemetry

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S22

### Business Value

Govern spend with visibility

### Persona

FinOps lead

### Assumptions / Risks

Budget alerts per team/model; Incorrect attribution

**Story** *As a FinOps lead, I want per-team/model/token cost views so spend is governed.*

**Non-Functional** Cost Reliability

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** cost export is available

**When** we build dashboards and budgets

**Then** alerts fire at thresholds

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Build cost attribution ETL
- ☐ Create dashboards per team/model
- ☐ Configure budget alerts



## S25 — Closed-beta rollout

### Epic / Feature

Beta Rollout

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S17–S24

### Business Value

Validate value with real users

### Persona

product owner

### Assumptions / Risks

Feedback loop; access list; Pilot bias / small sample

**Story** *As a product owner, I want a small pilot cohort so we validate value safely.*

### Non-Functional

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a pilot cohort is selected

**When** we enable access and collect feedback

**Then** weekly evaluation deltas are reported

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Create pilot access list
- ☐ Set feedback channels
- ☐ Publish weekly eval deltas

## S26 — Safety review & sign-off

### Epic / Feature

Beta Rollout

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S25

### Business Value

Formal go-live approval

### Persona

compliance officer

### Assumptions / Risks

Complete checklists; residual risks accepted; Late findings delay launch

**Story** *As a compliance officer, I want final privacy/security review so we can go live.*

### Non-Functional

Security

Privacy

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** all checklists are ready

**When** we obtain approvals from security/privacy

**Then** a signed go-live record exists

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.

**Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Run final security/privacy checklist
- ☐ Record residual risks
- ☐ Capture sign-off record

## S27 — SLOs & error budgets

### Epic / Feature

GA Readiness

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S23

### Business Value

Reliability targets set and enforced

### Persona

SRE

### Assumptions / Risks

Define SLOs and runbooks; Unrealistic SLOs

**Story** *As an SRE, I want latency/availability SLOs and error budgets so reliability is managed.*

### Non-Functional

Reliability

Performance

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** baseline data is available

**When** we publish SLOs and error budgets

**Then** alerts and runbooks are live

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Set latency/availability SLOs
- ☐ Create alert runbooks
- ☐ Announce error budgets

## S28 — Incident response & red-teaming

**Epic / Feature**

GA Readiness

**Priority / Estimate**

Priority: Must    SP: TBD

**Dependencies**

S27

**Business Value**

Operational readiness validated

**Persona**

security lead

**Assumptions / Risks**

Run game-day; fix gaps; Uncovered gaps at go-live

**Story** *As a security lead, I want runbooks and a red-team exercise so we're operationally ready.*

**Non-Functional**

Security

Reliability

**Acceptance Criteria (BDD)**

**Scenario** Happy path

**Given** on-call and runbooks are in place

**When** we run a red-team/game-day

**Then** findings are addressed and recorded

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Schedule game-day
- ☐ Execute playbooks
- ☐ Track and close findings

## S29 — Feature flags & kill-switch

### Epic / Feature

GA Readiness

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S27

### Business Value

Rapid rollback capability

### Persona

platform engineer

### Assumptions / Risks

Per-capability flags; audit logs; Broken toggles

**Story** *As a platform engineer, I want per-capability flags and a kill-switch so risky behavior can be rolled back instantly.*

### Non-Functional

Reliability

Security

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a flag service is ready

**When** we add flags and a kill-switch

**Then** prod/staging toggles are verified

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Add per-capability flags
- ☐ Implement kill-switch
- ☐ Verify toggles in staging/prod

## S30 — Prompt/model A/B

### Epic / Feature

Continuous Improvement

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S18, S22

### Business Value

Faster iteration with evidence

### Persona

product analyst

### Assumptions / Risks

Experiment registry; stat-sig criteria; Noisy results

**Story** *As a product analyst, I want A/B infra for prompts/models so we ship measurable gains.*

**Non-Functional** Performance Reliability Cost

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** traffic and segmentation are ready

**When** we enable prompt/model experiments

**Then** stat-sig criteria are documented and used

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

## Tasks

- ☐ Create experiment registry
- ☐ Route traffic to variants
- ☐ Define stat-sig criteria

## S31 — Caching & retrieval optimization

### Epic / Feature

Continuous Improvement

### Priority / Estimate

Priority: Must SP: TBD

### Dependencies

S17, S22

### Business Value

Lower latency and cost

### Persona

platform engineer

### Assumptions / Risks

Cache invalidation rules; Stale cache responses

**Story** *As a platform engineer, I want response and embedding-cache layers so cost/latency drop.*

### Non-Functional

Performance

Cost

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a cache service is available

**When** we add response and embedding caches

**Then** p95 latency and cost targets are met

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set. **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Add response cache
- ☐ Add embedding cache
- ☐ Set invalidation rules and metrics

## S32 — Model portfolio & fallback

### Epic / Feature

Continuous Improvement

### Priority / Estimate

Priority: Must    SP: TBD

### Dependencies

S30, S31

### Business Value

Balance UX and cost with resilience

### Persona

architect

### Assumptions / Risks

Router policies and failover paths; Misroutes / degraded quality

**Story** *As an architect, I want tiered models (cheap/fast vs accurate) with fallbacks so UX and cost are balanced.*

### Non-Functional

Performance

Reliability

Cost

### Acceptance Criteria (BDD)

**Scenario** Happy path

**Given** a model router is in place

**When** we configure routing policies and failover

**Then** failover is tested and documented

**Definition of Ready:** Persona clear; AC drafted; Dependencies known; Estimate set.    **Definition of**

**Done:** All ACs pass; Tests green; Security/all checks; Docs updated; Deployed/flagged.

### Tasks

- ☐ Define routing rules
- ☐ Implement failover paths
- ☐ Test failover on golden cases