# AI Infrastructure Platform – Planning & Delivery Document (Draft)

**Audience:** Executive stakeholders, Platform leadership, Architecture review
**Author:** Raj / Bruno
**Purpose:** Establish a clear, phased plan to build a governed AI infrastructure that demonstrably improves DHCS outcomes, while enabling faster, safer delivery and recovery.

## 1. Problem Statement

DHCS teams are increasingly expected to apply AI to complex, high-risk workflows (e.g., intake processing, grading, summarization, review support). Today, AI efforts risk becoming:

- Fragmented and one-off
- Difficult to govern and audit
- Hard to evaluate for quality and safety
- Slow to scale across teams

At the same time, leadership needs **proof** that AI:

- Improves efficiency and outcomes
- Is reliable and recoverable when failures occur
- Can be scaled responsibly without increasing risk

## 2. Strategy Overview (AI-First Framing)

### Core Strategy

Build a shared, governed AI infrastructure first. Use delivery speed, reliability, and recovery metrics to prove value. Instead of optimizing for developer productivity in isolation, this platform:

- Enables **safe, repeatable AI capabilities**
- Makes **quality, auditability, and failure recovery explicit**
- Allows developer productivity to emerge as a measurable outcome

# 3. Guiding Principles

1. **AI is advisory, not autonomous**
   Human oversight is preserved for all critical decisions.
2. **Governance is built-in, not bolted on**
   Every AI workload inherits auditability, access control, and quality checks.
3. **Failure is expected and designed for**
   AI failures degrade safely and recover quickly.
4. **State-owned and sustainable**
   Architecture avoids lock-in and supports long-term DHCS ownership.
5. **Metrics over anecdotes**
   Success is demonstrated through measurable outcomes, not demos alone.

# 4. Assumptions / Needs Confirmation

The following assumptions must be validated early:

1. **AI Decision Boundaries**
   a. AI outputs are advisory unless explicitly approved otherwise.
2. **Data Classification Policy**
   a. Clear definitions for PHI, PII, sensitive internal, and public data exist or will be finalized.
3. **Model Hosting Constraints**
   a. Whether inference must be fully state-controlled or can include approved external services.
4. **Failure Tolerance**
   a. Which workflows must continue operating if AI is unavailable.
5. **Initial Lighthouse Use-Cases**
   a. 2–3 priority workflows identified to demonstrate early AI impact.
6. **Compliance Authority**
   a. Defined ownership for approving AI policies, evaluation thresholds, and release gates.

# 5. Key Risks and Mitigations

## Risk 1: AI quality issues reduce trust

**Mitigation**

- Mandatory offline evaluation before release

- Human override tracking
- Continuous quality monitoring

### Risk 2: AI failures block critical workflows

**Mitigation**

- Designed degradation paths (fallback to rules/manual)
- Explicit AI disable switches
- Clear on-call ownership and runbooks

### Risk 3: Platform seen as slowing teams down

**Mitigation**

- Golden paths and templates
- Pre-approved compliance patterns
- Early focus on high-value, visible wins

### Risk 4: Governance becomes manual or inconsistent

**Mitigation**

- Policy enforcement through infrastructure
- Automated audit logging
- Centralized documentation and standards

### Risk 5: Long-term dependency on specific vendors

**Mitigation**

- Clear abstraction layers
- Documented exit strategies
- Emphasis on open interfaces and standards

## 6. Phased Delivery Plan

### Phase 0 – Alignment & Foundations

**Objective:** Establish clarity, scope, and measurable success criteria.

### Deliverables

- Platform charter (what the platform is / is not)
- AI use-case inventory and prioritization
- Initial success metrics definition
- Governance boundaries and assumptions documented

### Goals

- Executive alignment on scope and expectations
- Clear definition of “success” for AI adoption

### Metrics

- Agreement on prioritized use-cases
- Approved platform charter
- Baseline metrics captured

## Phase 1 – Core AI Infrastructure (Highest Priority)

**Objective:** Make safe, governed AI execution possible.

### Deliverables

- Standard AI runtime for:
    - Prompt orchestration
    - Retrieval (RAG)
    - Multi-step workflows
- Built-in guardrails:
    - Structured outputs
    - Policy enforcement
    - Audit logging
- Human-in-the-loop checkpoints

### Goals

- All AI features run through a single governed runtime
- No ad-hoc AI logic in production systems

### Metrics

- % of AI workloads using standard runtime
- Audit completeness for AI executions
- Successful human override paths tested

## Phase 2 – AI Quality, Safety & Recovery

**Objective:** Ensure AI systems are trustworthy and resilient.

### Deliverables

- Offline evaluation harness (golden datasets, rubrics)
- Online monitoring:
  - Error types
  - Drift indicators
  - Human overrides
- Failure and recovery patterns:
  - Safe degradation
  - Clear ownership
  - Incident runbooks

### Goals

- AI regressions detected before production
- Failures recover quickly without blocking operations

### Metrics

- Evaluation pass rate before release
- Change failure rate for AI features
- Mean Time to Recovery (MTTR)

## Phase 3 – Delivery & Scaling System

**Objective:** Prove AI accelerates delivery safely.

### Deliverables

- Golden-path templates for AI-enabled services
- CI/CD with integrated eval and policy gates
- Standard observability and dashboards

### Goals

- Teams ship AI features faster with less bespoke work
- Reduced operational burden per new use-case

### Metrics

- Lead time to ship AI features
- Deployment frequency
- Reduction in platform team intervention

## Phase 4 – Adoption, Transparency & ROI

**Objective:** Make value visible and sustainable.

### Deliverables

- Platform-level dashboards:
  - Usage
  - Quality trends
  - Cost vs impact
- AI Hub documentation and onboarding guides
- Ongoing support and training model

### Goals

- Leadership has continuous visibility into AI impact
- Platform adoption grows without increased risk

### Metrics

- Adoption across teams
- Human time saved (validated samples)
- Quality and cost trends over time

# 7. How This Proves AI Value

This approach demonstrates success by showing that:

- AI reduces manual effort **without reducing accountability**
- AI failures are **contained, visible, and recoverable**

- New AI capabilities can be delivered **faster and more safely over time**

Developer productivity, shipping speed, and recovery metrics are **evidence** that the AI platform is working—not the initial promise.

## 8. Next Steps

1. Review assumptions with leadership (Bruno + stakeholders)
2. Confirm initial lighthouse use-cases
3. Approve Phase 0 deliverables and timeline
4. Begin Phase 1 implementation