

Productionize Agent Canvas

Based on the excellent MLOps Canvas stack here: MLOps Stack Canvas

This canvas serves as a comprehensive project planning tool for agentic AI implementations, organized into three logical phases:

DESIGN & ARCHITECTURE	Define the problem, design the solution, and plan the cognitive architecture
IMPLEMENTATION & CONSTRAINTS	Address practical limitations and build the system
VALIDATION & OPERATIONS	Test the system and plan for ongoing success

For questions, please reach out to:

[the author, Mihko Chandrasekhar](#)
[AI/ML DevRel, Richmond AI/ate](#)
[AI/ML DevRel, Pavel Duchovny](#)

The canvas helps teams systematically work through all aspects of an agentic AI project while avoiding redundancy and ensuring nothing critical is missed.

Project Name	Team Members
Date	

TIER 1: PRODUCT & SCALE			TIER 2: AGENT ARCHITECTURE		
Square 1 - BUSINESS CASE & SCALE PLANNING			Square 3 - ROBUST AGENT ARCHITECTURE		
From POC validation to production business case			From simple POC agent to production-grade system		
Proven Value	What did the POC validate about user adoption and business impact?		Workflow orchestration	How will complex workflows be managed at scale?	
Business Metrics	What production KPIs will measure ongoing success?		Multi-Agent Coordination	How will specialized agents work together?	
Scale Requirements	How many users, interactions, and usage patterns will production need to support?		Fault Tolerance	How will the system handle failures gracefully?	
Growth Planning	How will usage scale over time and what capabilities will be needed?		Update Rollouts	How will improvements be deployed safely?	

Square 2 - PRODUCTION REQUIREMENTS & CONSTRAINTS			Square 4 - PRODUCTION MEMORY & CONTEXT		
Production constraints the POC didn't address			From POC memory to production-scale context management		
Performance Requirements	Response time, availability, throughput standards?		Memory Architecture	What types of memory at what scale?	
Reliability Standards	How does the system recover from failures?		Context Persistence	How will context be stored and retrieved?	
Budget Constraints	What are the cost limits and optimization targets for sustainable operation?		Cross-Session Continuity	How will agents maintain context across interactions?	
Security Needs	Cost limits and optimization targets?		Memory Lifecycle Management	How will memory be cleaned and optimized?	

TIER 4:
PROJECT METADATA

PROJECT METADATA & GOVERNANCE		
Project Team	What production-focused team structure and roles are needed?	
Implementation Timeline	What are the key phases and milestones for production rollout?	
Resource Planning	What team and infrastructure costs are required for production?	
Success Metrics	What are the production readiness criteria and ongoing success measures?	
Risk Management	What are the key production-specific risks and mitigation strategies?	
Stakeholder Alignment	Who are the key decision makers and how will they be engaged?	

TIER 3: DATA INFRASTRUCTURE			TIER 4: AGENT OPERATIONS			TIER 5: HARDENING & OPERATIONS		
Square 5 - DATA ARCHITECTURE & MANAGEMENT			Square 7 - MODEL STRATEGY & OPTIMIZATION			Square 9 - SECURITY & COMPLIANCE		
From POC data enhancement to production data operations			Managing external models and fine-tuning in production			Production security and regulatory compliance		
Platform Architecture	How will application, vector, and memory data be unified?		Provider Selection	Which providers and models for different use cases?		Security Implementation	Authentication, authorization, encryption?	
Data Pipelines	How will data be ingested, processed, and updated?		Fine-tuning Approach	When and how to create custom models?		Access Control	User and system access management?	
Quality Assurance	How will accuracy and freshness be ensured?		Routing Logic	How to route between base and fine-tuned models?		Compliance Framework	What regulations must be addressed?	
Knowledge Governance	How will information be maintained and governed?		Cost Controls	How to balance performance and cost?		Audit & Logging	What audit capabilities are needed?	
Square 6 - KNOWLEDGE BASE & OPERATIONS			Square 8 - API MANAGEMENT & MONITORING			Square 10 - USER EXPERIENCE & ADOPTION		
Continuous data improvement and operations			Managing external APIs and tracking fine-tuned model performance			From POC user testing to production user management		
Update Strategy	How will the knowledge base evolve?		API Configuration	How to manage multiple model endpoints?		Workflow Integration	How will the agent integrate into workflows?	
Embedding Approach	What embedding models for what content?		Performance Tracking	How to compare base vs fine-tuned models?		Adoption Strategy	How will usage be driven and sustained?	
Retrieval Optimization	How will search relevance be maximized?		Fine-tuning Pipeline	How to collect data and retrain?		Support Systems	How will users get help when needed?	
Operational Monitoring	How will data flows be monitored and optimized?		Version Control	How to handle multiple fine-tuned versions?		Feedback Integration	How will user input drive improvements?	

Square 11 - CONTINUOUS IMPROVEMENT & GOVERNANCE		
Production governance and long-term sustainability		
Operational Procedures	How will the system be maintained?	
Quality Assurance	How will quality be maintained over time?	
Cost Management	How will costs be controlled long-term?	
Continuity Planning	How will operational knowledge be preserved?	