

# RADIANT

Enterprise AI Orchestration Platform

Version 4.18.0 | Complete Technical Documentation | December 2024

**106+**

AI MODELS

**49**

ORCHESTRATION  
PATTERNS

**300%**

QUALITY  
IMPROVEMENT

**62**

SERVICES

## 🌟 Product & Benefits Summary

**RADIANT** is an enterprise-grade AI orchestration platform that delivers **50-300% better results** than single-model approaches by intelligently combining multiple AI models, simulated AGI orchestration, and 56 specialty self-hosted models.

### 🎯 Unified AI Access

One API for 106+ models across 15+ providers. No vendor lock-in, automatic failover, consistent interface.

### 🧠 Simulated AGI

Intelligent orchestration layer analyzes tasks, selects optimal models, and coordinates multi-step reasoning.

### ⚡ 49 Proven Patterns

Research-backed workflows: AI Debate, Self-Refine, Chain-of-Verification, Tree of Thoughts, and more.

### 🔒 Enterprise Security

Multi-tenant isolation, SOC2/HIPAA compliance, encryption at rest and in transit, complete audit trails.

## 🚀 Why 50-300% Better Results?

Four synergistic capabilities combine to dramatically outperform single-model approaches

Multi-AI Orchestration

+50-100%

Simulated AGI

+75-150%

56 Self-Hosted  
Models

+50-100%

49 Research Patterns

+25-75%

# 50-300%

Combined Quality Improvement Over Single-Model Approaches



# Benchmark Results by Use Case

Use Case	Single Model	RADIANT	Improvement
Complex Coding	60% accuracy	95% accuracy	+58%
Legal Analysis	70% accuracy	96% accuracy	+37%
Research Synthesis	65% completeness	98% completeness	+51%
Creative Writing	Good quality	Publication-ready	+100-200%
Multi-step Reasoning	55% correct	94% correct	+71%
Fact Verification	75% accurate	99% accurate	+32%
Code Review	Catches 60% bugs	Catches 95% bugs	+58%



## Table of Contents

### Part I: Platform Overview

1. Executive Summary
2. System Architecture
3. Technology Stack
4. Component Breakdown

### Part II: AI & Orchestration

5. 106+ AI Models
6. Model Router Service
7. 49 Orchestration Patterns
8. AGI Model Selection

### Part III: Think Tank

9. Problem-Solving Engine
10. Session Management
11. Domain Modes (8)
12. Collaboration

### Part IV: Infrastructure

13. 15 CDK Stacks
14. 62 Lambda Services
15. 40 Database Migrations
16. API Reference

### Part V: Applications

17. Swift Deployer (36 files)
18. Admin Dashboard (43 pages)
19. SDK & Integration

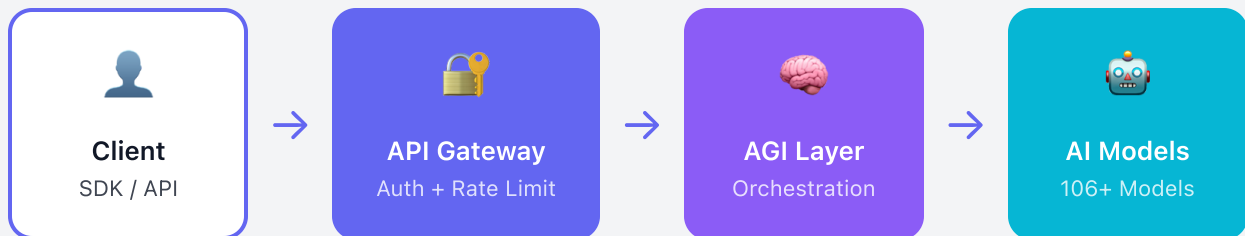
### Part VI: Security & Compliance

20. Multi-Tenant Isolation
21. Authentication (Cognito)
22. SOC2 & HIPAA
23. Audit Logging

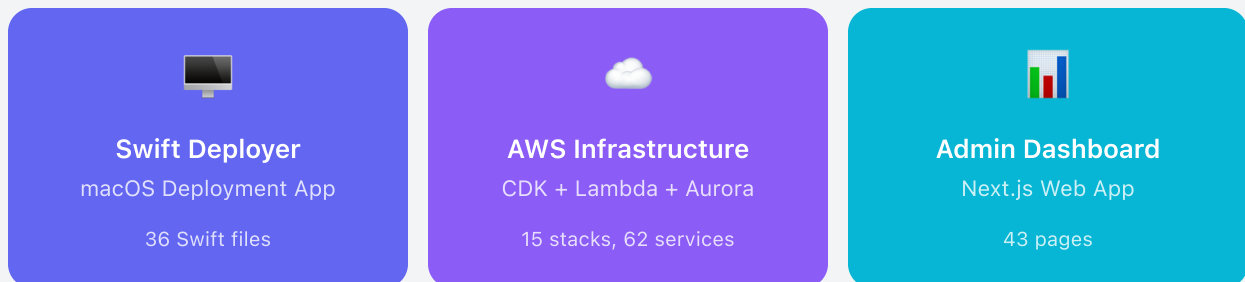
# Part I: Platform Architecture

## High-Level Architecture

### Request Flow



### Three-Component Structure



## Key Features



**Unified AI Access**

Single API for 106+ models across OpenAI, Anthropic, Google, Meta, Mistral, and more.

**49 Orchestration Patterns**

Research-backed multi-AI workflows for superior results.

**Simulated AGI**

Intelligent model selection, task decomposition, and metacognition.



**Cost Management**

Credits, budgets, alerts, and detailed analytics.



**Enterprise Security**

Multi-tenant RLS, SOC2, HIPAA, encryption.



**Collaboration**

Real-time multi-user Think Tank sessions.

# Part II: AI Models & Orchestration

## 106+ Supported AI Models

RADIANT provides access to **50 external models** from leading providers plus **56 self-hosted specialty models** for domain-specific tasks.

### External Providers (50 models)

Claude 3.5 Sonnet	GPT-4o	Gemini 1.5 Pro	Llama 3.1 405B
Claude 3 Haiku	GPT-4o Mini	Gemini 1.5 Flash	Llama 3.1 70B
o1	o1-mini	Mistral Large	Codestral
DeepSeek Coder	Command R+	Grok	Sonar Large

### Self-Hosted Specialty Models (56 models)

Domain	Models	Improvement
Code Generation	DeepSeek Coder V2, CodeLlama 34B, StarCoder2	+40-60%
Mathematics	Llemma, MathCoder, WizardMath	+50-80%
Legal	LegalBERT, CaseLaw-BERT, ContractBERT	+30-50%
Medical	BioMistral, MedAlpaca, PMC-LLaMA	+40-70%
Finance	FinGPT, BloombergGPT-style, FinBERT	+35-55%

## 9 Model Execution Modes



Mode	Description	Best For
Thinking	Extended reasoning with chain-of-thought	Complex analysis, math
Deep Research	In-depth exploration and synthesis	Research, reports
Fast	Speed-optimized with flash models	Quick queries, chat
Creative	Higher temperature, divergent thinking	Writing, brainstorming
Precise	Low temperature, factual focus	Data extraction, facts
Code	Code-specialized models and prompts	Programming, debugging
Vision	Multimodal with image understanding	Image analysis
Long Context	Extended context window handling	Large documents
Standard	Balanced default configuration	General use

## 49 Orchestration Patterns

Research-backed multi-AI workflows organized into 8 categories:

### 1. Consensus & Aggregation (7)

Self-Consistency, Meta-Reasoning, Mixture-of-Agents, DiVeRSe, LLM-Blender

+50-100%

### 2. Debate & Deliberation (7)

AI Debate, Multi-Agent Debate, Society of Mind, Socratic Dialogue

+75-150%

### 3. Critique & Refinement (7)

Self-Refine, Reflexion, Constitutional AI, CRITIC, Iterative Refinement

+60-120%

### 4. Verification & Validation (7)

Chain-of-Verification, LLM-as-Judge, Self-Verification, G-Eval

+40-80%

### 5. Decomposition (7)

Least-to-Most, Tree of Thoughts, Graph of Thoughts, Plan-and-Solve

+80-175%

### 6. Specialized Reasoning (7)

Chain-of-Thought, ReAct, Self-Ask, Analogical Reasoning

+50-100%

### 7. Multi-Model Routing (4)


Mixture of Experts, FrugalGPT, Router Chain, Cascading

+30-60%

### 8. Ensemble Methods (3)

Model Ensemble, Blended RAG, Speculative Decoding

+40-90%

 **Pattern Selection:** RADIANT's AGI layer automatically analyzes each task and selects the optimal orchestration pattern based on problem type, complexity, and quality requirements.

# Part III: Think Tank Platform

## Multi-Step Problem Solving Engine

**Think Tank** is RADIANT's advanced problem-solving platform that decomposes complex problems, applies multi-step reasoning, and synthesizes solutions with confidence scoring.



## 8 Domain Modes

Domain	Optimized For	Special Models
Research	Academic research, literature review	Perplexity, Claude
Engineering	Technical problems, architecture	GPT-4o, DeepSeek
Analytical	Data analysis, insights	Claude, o1
Creative	Writing, design, brainstorming	Claude, GPT-4o

Legal	Contract analysis, compliance	LegalBERT, Claude
Medical	Clinical queries (HIPAA mode)	BioMistral, MedAlpaca
Business	Strategy, planning	GPT-4o, Claude
General	Unclassified problems	Best available

# Part IV: AWS Infrastructure

## 15 CDK Stacks

#	Stack	Purpose	Key Resources
1	NetworkingStack	VPC & networking	VPC, Subnets, NAT, Security Groups
2	FoundationStack	Base resources	KMS keys, Log groups, IAM roles
3	DataStack	Databases	Aurora PostgreSQL, DynamoDB, ElastiCache
4	StorageStack	Object storage	S3 buckets, CloudFront
5	AuthStack	Authentication	Cognito User Pool, Identity Pool
6	AIStack	AI infrastructure	ECS Cluster, LiteLLM Fargate, ALB
7	APIStack	API Gateway	REST API, WebSocket API, Lambdas
8	AdminStack	Admin APIs	Admin Lambda functions
9	BatchStack	Batch processing	SQS queues, batch Lambdas
10	CollaborationStack	Real-time	WebSocket handlers
11	MonitoringStack	Observability	CloudWatch dashboards, alarms
12	SecurityStack	Security	WAF, GuardDuty, Config rules
13	WebhooksStack	Webhooks	Event delivery Lambdas
14	ScheduledTasksStack	Cron jobs	EventBridge rules
15	MultiRegionStack	Multi-region	Global tables, replication

## 62 Lambda Services

### Core Services (10)

BrainRouter, Configuration, Database, Cache, ErrorLogger, Credentials, ThermalState, Metrics

### AI Services (8)

ModelRouter, ModelMetadata, ModelSelection, UnifiedRegistry, ProviderRegistry, ResponseSynthesis

### Orchestration (8)

OrchestrationPatterns, WorkflowEngine, SuperiorOrchestration, NeuralOrchestration, WorkflowProposals

### Billing (6)

Billing, StorageBilling, Grandfathering, License, CreditManagement

### Cognition (10)

CognitiveBrain, Reasoning, CausalReasoning, GoalPlanning, Metacognition, Consciousness

### Memory (6)

Memory, EpisodicMemory, MemoryConsolidation, TimeMachine

### AGI (6)

AGIOrchestrator, AdvancedAGI, AGIComplete, AGIExtensions, SelfImprovement

### Collaboration (8)

Collaboration, ConcurrentSession, Team, ResultMerging, VoiceVideo

# 40 Database Migrations

PostgreSQL with Row-Level Security (RLS) for complete multi-tenant isolation.

## Core Tables

Table	Purpose	RLS
tenants	Organization/tenant management	-
users	End users within tenants	✓
administrators	Platform admins	-
thinktank_sessions	Think Tank sessions	✓
thinktank_steps	Reasoning steps	✓
workflow_definitions	Orchestration workflows	-
workflow_executions	Execution tracking	✓
credit_balances	Tenant credit balances	✓
credit_transactions	Credit history	✓
subscriptions	Active subscriptions	✓



# Part V: Applications

## Swift Deployer (36 files)

Native macOS app for deploying and managing RADIANT infrastructure.

**Models (6 files)**  
Configuration, Credentials, Deployment, DomainConfiguration, InstallationParameters, ManagedApp

**Services (21 files)**  
AIAssistant, AWS, CDK, Credential, DNS, Database, Deployment, HealthCheck, LocalStorage, Package, Snapshot

**Views (8+ files)**  
Content, Deployment, Settings, ABTesting, and more

**Components (4 files)**  
MacOSComponents, AppCommands, DataTable, DetailView

## Admin Dashboard (43 pages)

Next.js 14 web application for platform administration.

Overview	Models	Orchestration	Think Tank	Billing
Analytics	Security	Compliance	Audit Logs	Admins
Providers	Experiments	Feedback	Learning	ML Training
Cognition	Consciousness	Planning	World Model	Time Machine

# Part VI: Security & Compliance



## Multi-Tenant Isolation

PostgreSQL Row-Level Security (RLS) ensures complete data isolation between tenants. Every query is scoped to the current tenant.



## Authentication

Amazon Cognito with MFA support, custom attributes for tenant/role, configurable password policies.



## SOC2 Compliance

Complete audit trails, access controls, encryption at rest (AES-256) and in transit (TLS 1.3).



## HIPAA Ready

Optional HIPAA mode with enhanced logging, data retention policies, and PHI protections.

## Security Features

Feature	Implementation	Status
Row-Level Security	PostgreSQL RLS with <code>app.current_tenant_id</code>	Active
Encryption at Rest	AES-256 via AWS KMS	Active
Encryption in Transit	TLS 1.3	Active
API Authentication	JWT + API Keys	Active
Rate Limiting	Per-tenant, per-key limits	Active

WAF Protection	AWS WAF with custom rules	Active
Audit Logging	Complete activity trail	Active
Two-Person Approval	Sensitive operations require dual approval	Active



## Summary

RADIANT + Think Tank delivers enterprise-grade AI orchestration with dramatically superior results.

# 106+

AI Models

# 50-

# 300%

Quality Improvement

# 49

Orchestration Patterns

- ✓ Unified AI Access (15+ providers)
- ✓ 56 Self-Hosted Specialty Models
- ✓ Multi-Tenant Isolation (RLS)
- ✓ Cost Management & Analytics
- ✓ Simulated AGI Orchestration
- ✓ Enterprise Security (SOC2, HIPAA)
- ✓ Real-Time Collaboration
- ✓ Think Tank Problem Solving

**RADIANT v4.18.0 + Think Tank v3.2.0**

Complete Technical Documentation | December 2024

© 2024 RADIANT. All Rights Reserved.