

# RADIANT Platform - Administrator Guide

RADIANT Team

2025-12-25

## Contents

<b>RADIANT Platform - Administrator Guide</b>	<b>3</b>
Table of Contents . . . . .	3
1. Introduction . . . . .	3
1.1 What is RADIANT? . . . . .	3
1.2 Administrator Roles . . . . .	4
1.3 Key Concepts . . . . .	4
2. Accessing the Admin Dashboard . . . . .	5
2.1 URL and Login . . . . .	5
2.2 First Login . . . . .	5
2.3 Session Management . . . . .	6
2.4 Password Requirements . . . . .	6
3. Dashboard Overview . . . . .	6
3.1 Main Dashboard . . . . .	6
3.2 Navigation Menu . . . . .	7
4. Tenant Management . . . . .	7
4.1 Viewing Tenants . . . . .	7
4.2 Creating a Tenant . . . . .	7
4.3 Tenant Details . . . . .	8
4.4 Tenant Actions . . . . .	8
4.5 Data Isolation . . . . .	8
5. User & Administrator Management . . . . .	9
5.1 Administrator Roles . . . . .	9
5.2 Managing Administrators . . . . .	9
5.3 Viewing Tenant Users . . . . .	9
5.4 User Actions . . . . .	10
6. AI Model Configuration . . . . .	10
6.1 Model Registry . . . . .	10
6.2 Model Categories . . . . .	10
6.3 Model Configuration . . . . .	11
6.4 Self-Hosted Models . . . . .	12
6.5 Thermal States (Self-Hosted) . . . . .	12
7. Provider Management . . . . .	12
7.1 External Providers . . . . .	12
7.2 Adding Provider Credentials . . . . .	13

7.3 Provider Health Monitoring . . . . .	13
7.4 Fallback Configuration . . . . .	13
8. Billing & Subscriptions . . . . .	13
8.1 Subscription Tiers . . . . .	13
8.2 Credit System . . . . .	14
8.3 Managing Subscriptions . . . . .	14
8.4 Usage Reports . . . . .	14
8.5 Billing Alerts . . . . .	14
9. Storage Management . . . . .	15
9.1 Storage Overview . . . . .	15
9.2 Storage Tiers . . . . .	15
9.3 File Management . . . . .	15
10. Orchestration & Neural Engine . . . . .	16
10.1 Brain Router . . . . .	16
10.2 Neural Patterns . . . . .	16
10.3 Workflow Templates . . . . .	16
11. Localization . . . . .	16
11.1 Translation Management . . . . .	16
11.2 Supported Languages . . . . .	16
11.3 AI Translation . . . . .	17
12. Configuration Management . . . . .	17
12.1 System Configuration . . . . .	17
12.2 Tenant Overrides . . . . .	17
12.3 SSM Parameters . . . . .	17
13. Security & Compliance . . . . .	18
13.1 Security Dashboard . . . . .	18
13.2 Anomaly Detection . . . . .	18
13.3 Compliance Reports . . . . .	18
13.4 Generating Reports . . . . .	18
14. Cost Analytics . . . . .	19
14.1 Cost Dashboard . . . . .	19
14.2 Cost Alerts . . . . .	19
14.3 Cost Optimization . . . . .	19
15. A/B Testing & Experiments . . . . .	20
15.1 Experiment Dashboard . . . . .	20
15.2 Creating an Experiment . . . . .	20
15.3 Statistical Analysis . . . . .	20
16. Audit & Monitoring . . . . .	20
16.1 Audit Logs . . . . .	20
16.2 Log Filtering . . . . .	21
16.3 Log Export . . . . .	21
16.4 Real-Time Monitoring . . . . .	21
17. Database Migrations . . . . .	21
17.1 Migration Workflow . . . . .	21
17.2 Pending Migrations . . . . .	21
17.3 Approving Migrations . . . . .	22
18. API Management . . . . .	22
18.1 API Keys . . . . .	22

18.2 Rate Limiting . . . . .	22
18.3 Webhooks . . . . .	22
19. Troubleshooting . . . . .	23
19.1 Common Issues . . . . .	23
19.2 Support Resources . . . . .	23
19.3 Log Locations . . . . .	23
Appendix: Quick Reference . . . . .	24
Keyboard Shortcuts . . . . .	24
Status Indicators . . . . .	24

## RADIANT Platform - Administrator Guide

**Complete guide for managing the RADIANT AI Platform via the Admin Dashboard**

Version: 4.18.1 | Last Updated: December 2024

---

### Table of Contents

1. Introduction	
2. Accessing the Admin Dashboard	
3. Dashboard Overview	
4. Tenant Management	
5. User & Administrator Management	
6. AI Model Configuration	
7. Provider Management	
8. Billing & Subscriptions	
9. Storage Management	
10. Orchestration & Neural Engine	
11. Localization	
12. Configuration Management	
13. Security & Compliance	
14. Cost Analytics	
15. A/B Testing & Experiments	
16. Audit & Monitoring	
17. Database Migrations	
18. API Management	
19. Troubleshooting	

---

## 1. Introduction

### 1.1 What is RADIANT?

RADIANT is a multi-tenant AWS SaaS platform providing unified access to 106+ AI models through:

- **50 External Provider Models:** OpenAI, Anthropic, Google, xAI, DeepSeek, and more
- **56 Self-Hosted Models:** Running on AWS SageMaker for cost control and privacy
- **Intelligent Routing:** Brain router for optimal model selection
- **Neural Engine:** Personalization learning from user interactions

## 1.2 Administrator Roles

Role	Permissions	Use Case
<b>Super Admin</b>	Full access to all features	Platform owner
<b>Admin</b>	Tenant management, billing, models	Operations team
<b>Operator</b>	Read access, limited actions	Support team
<b>Auditor</b>	Read-only access to logs	Compliance team

**Role Details Super Admin** - The highest privilege level with unrestricted access: - Create and delete tenants - Manage all administrators - Access all billing and financial data - Modify system-wide configuration - Approve production database migrations - Impersonate any tenant for debugging - Access compliance and audit reports - Typically limited to 1-3 people (CTO, lead engineer)

**Admin** - Day-to-day operations management: - Create and modify tenants (cannot delete) - Manage users within tenants - Configure AI models and providers - View billing data (cannot modify pricing) - Monitor system health - Cannot access other admin accounts - Typically assigned to operations team members

**Operator** - Limited support and monitoring: - View tenant information (read-only) - View user issues and support tickets - Monitor system health dashboards - Cannot modify any configuration - Cannot access billing or sensitive data - Typically assigned to support staff

**Auditor** - Compliance and security review: - Full read access to audit logs - Access to compliance reports - Cannot modify anything - Cannot view sensitive data (API keys, passwords) - Access is logged for compliance - Typically assigned to compliance officers or external auditors

## 1.3 Key Concepts

Concept	Description
<b>Tenant</b>	Organization with isolated data
<b>User</b>	End-user within a tenant
<b>Subscription</b>	Billing tier (1-7)
<b>Credits</b>	Currency for AI usage
<b>API Key</b>	Authentication for API access
<b>App</b>	Consumer application (Think Tank, etc.)

**Tenant Architecture Explained** A **Tenant** represents a complete organization using RADIANT. Each tenant has:

- **Complete Data Isolation:** All data is stored with tenant IDs and protected by PostgreSQL Row-Level Security (RLS). One tenant can never access another tenant's data, even if there's a bug in application code.

- **Separate Billing:** Each tenant has its own subscription, credit balance, and usage tracking. Costs are attributed to the correct tenant automatically.
- **Custom Configuration:** Tenants can customize model access, rate limits, and feature flags without affecting other tenants.
- **User Management:** Each tenant manages their own users, roles, and permissions independently.

**User vs Administrator** Users are end-users who interact with RADIANT-powered applications like Think Tank. They: - Sign up and log in via Cognito - Use AI models through the API or applications - Have credits deducted for usage - Cannot access the Admin Dashboard

**Administrators** manage the RADIANT platform itself. They: - Access the Admin Dashboard - Manage tenants, users, and billing - Configure AI models and providers - Have no credits (administrative access is separate)

**Credit System Explained** Credits are RADIANT's universal currency for AI usage:

- **1 credit = \$0.01 USD** (configurable per deployment)
- Different models cost different amounts based on their API pricing
- Credits are deducted in real-time as requests complete
- Tenants can purchase credits or receive them through subscriptions
- Credits can be tracked, audited, and reported on

**Example Credit Costs:** | Model | Cost per 1K tokens | | GPT-4o | 5 credits input, 15 credits output | | GPT-4o-mini | 0.5 credits input, 1.5 credits output | | Claude 3.5 Sonnet | 3 credits input, 15 credits output | | Self-hosted Llama | 0.2 credits (all) |

**API Key Types** RADIANT supports multiple API key types:

- **User API Keys:** Tied to a specific user, inherit user's permissions
- **Service API Keys:** For server-to-server communication, not tied to a user
- **Admin API Keys:** For administrative operations, require elevated permissions
- **Scoped Keys:** Limited to specific models, endpoints, or rate limits

## 2. Accessing the Admin Dashboard

### 2.1 URL and Login

1. Navigate to: <https://admin.your-domain.com>
2. Enter your email address
3. Enter your password
4. Complete MFA verification (required)

### 2.2 First Login

On first login:

1. You'll receive a temporary password via email
2. Enter the temporary password

- 3. Set a new password (12+ characters, mixed case, numbers, symbols)
- 4. Set up MFA using an authenticator app
- 5. You'll be redirected to the dashboard

2.3 Session Management

Setting	Value
Session Duration	8 hours
Idle Timeout	30 minutes
Concurrent Sessions	3 maximum
Remember Device	30 days

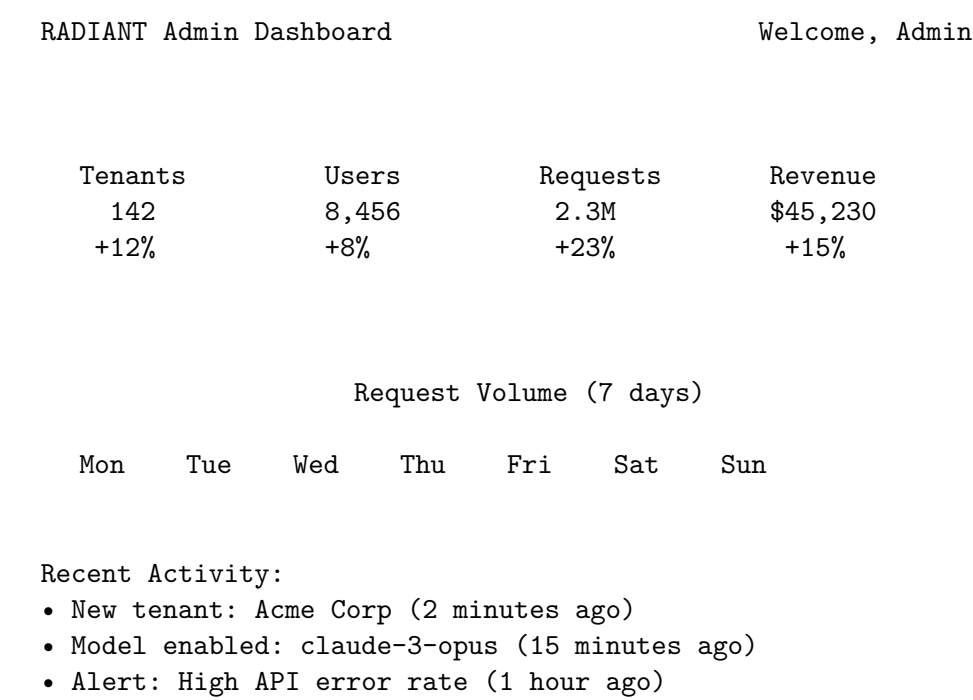
2.4 Password Requirements

- Minimum 12 characters
- At least one uppercase letter
- At least one lowercase letter
- At least one number
- At least one special character
- Cannot reuse last 10 passwords

3. Dashboard Overview

3.1 Main Dashboard

The dashboard displays key metrics at a glance:



### 3.2 Navigation Menu

Section	Description
<b>Dashboard</b>	Overview and metrics
<b>Tenants</b>	Tenant management
<b>Users</b>	User management
<b>Models</b>	AI model configuration
<b>Providers</b>	Provider management
<b>Billing</b>	Subscriptions and credits
<b>Storage</b>	Storage usage
<b>Orchestration</b>	Neural engine settings
<b>Localization</b>	Translation management
<b>Configuration</b>	System settings
<b>Security</b>	Security monitoring
<b>Compliance</b>	Compliance reports
<b>Experiments</b>	A/B testing
<b>Cost</b>	Cost analytics
<b>Audit</b>	Audit logs
<b>Migrations</b>	Database migrations
<b>Notifications</b>	System alerts
<b>Settings</b>	Personal settings

## 4. Tenant Management

### 4.1 Viewing Tenants

Navigate to **Tenants** to see all organizations:

Column	Description
<b>Name</b>	Organization name
<b>Plan</b>	Subscription tier
<b>Users</b>	User count
<b>Status</b>	Active/Suspended/Trial
<b>Created</b>	Creation date
<b>Last Active</b>	Last API call

### 4.2 Creating a Tenant

1. Click “+ New Tenant”
2. Fill in required fields:
  - **Name:** Organization name
  - **Slug:** URL-friendly identifier

- **Plan:** Initial subscription tier
  - **Admin Email:** Primary admin email
3. Configure optional settings:
- Custom domain
  - Branding settings
  - Feature flags
4. Click **“Create Tenant”**

4.3 Tenant Details

View comprehensive tenant information:

Tenant: Acme Corporation

Overview	Users	Billing	Settings
----------	-------	---------	----------

Tenant ID:	tn_abc123xyz
Status:	Active
Plan:	Professional (Tier 4)
Created:	2024-01-15
Last Active:	2 minutes ago

Usage This Month:	
API Requests:	145,234
Tokens Used:	12.5M
Storage:	2.3 GB
Credits Used:	\$1,234.56

[Edit]   [Suspend]   [Delete]   [Impersonate]

4.4 Tenant Actions

Action	Description	Permission
<b>Edit</b>	Modify tenant settings	Admin
<b>Suspend</b>	Temporarily disable	Admin
<b>Delete</b>	Permanently remove	Super Admin
<b>Impersonate</b>	Login as tenant admin	Super Admin
<b>Export</b>	Export tenant data	Admin

4.5 Data Isolation

Each tenant has complete data isolation:

- Separate database rows with RLS
- Unique API keys
- Isolated storage buckets
- Independent usage tracking

---

## 5. User & Administrator Management

### 5.1 Administrator Roles

Role	Dashboard Access	API Access	Billing	Audit
<b>Super Admin</b>	Full	Full	Full	Full
<b>Admin</b>	Full	Full	Read	Read
<b>Operator</b>	Read	Read	None	Read
<b>Auditor</b>	Logs only	None	None	Full

### 5.2 Managing Administrators

Navigate to **Administrators** to:

1. **Invite New Admin:**
  - Click “+ Invite Administrator”
  - Enter email address
  - Select role
  - Click “Send Invitation”
2. **Modify Admin:**
  - Click on administrator row
  - Edit role or permissions
  - Click “Save Changes”
3. **Remove Admin:**
  - Click “Remove” button
  - Confirm removal
  - Admin’s sessions are invalidated immediately

### 5.3 Viewing Tenant Users

Navigate to **Tenants** → [Tenant] → **Users** to see:

Field	Description
<b>Email</b>	User email
<b>Name</b>	Display name
<b>Role</b>	Tenant role
<b>Status</b>	Active/Invited/Disabled
<b>Last Login</b>	Last authentication
<b>API Keys</b>	Number of active keys

5.4 User Actions

Action	Description
Reset Password	Send password reset email
Disable	Prevent login
Enable	Restore access
Delete	Remove user data
View Sessions	See active sessions

6. AI Model Configuration

6.1 Model Registry

Navigate to **Models** to see all available models:

AI Models

106 Total

Filter: [All ] Category: [All ] Status: [Enabled ]

Model	Provider	Category	Tier	Status
gpt-4o	OpenAI	Chat	1	Enabled
gpt-4o-mini	OpenAI	Chat	1	Enabled
claude-3-opus	Anthropic	Chat	2	Enabled
claude-3-sonnet	Anthropic	Chat	1	Enabled
gemini-pro	Google	Chat	1	Enabled
llama-3.1-70b	Self-Host	Chat	3	Enabled
whisper-large	Self-Host	Audio	3	Disabled

[+ Add Model]

[Import Models]

[Export Config]

6.2 Model Categories

Category	Description	Example Models
Chat/LLM	Text generation	GPT-4o, Claude 3, Gemini
Embedding	Vector embeddings	text-embedding-3-large
Vision	Image understanding	GPT-4V, Claude Vision
Audio	Speech-to-text	Whisper, Deepgram
Image	Image generation	DALL-E 3, Stable Diffusion
Code	Code generation	Codestral, DeepSeek Coder
Scientific	Research models	BioGPT, ChemLLM

**Category Details Chat/LLM (Large Language Models):** The core of RADIANT. These models handle conversational AI, content generation, summarization, and general-purpose text tasks. They're the most commonly used and include flagship models from OpenAI, Anthropic, Google, and open-source alternatives.

**Embedding Models:** Convert text into numerical vectors for semantic search, similarity matching, and retrieval-augmented generation (RAG). Essential for building knowledge bases and search functionality. Vectors are typically 1536-3072 dimensions.

**Vision Models:** Analyze images, extract text (OCR), describe visual content, and answer questions about images. Increasingly important for document processing, accessibility, and multimodal applications.

**Audio Models:** Transcribe speech to text, translate audio, and identify speakers. Whisper is the most popular, offering excellent accuracy across 99 languages. Used for meeting transcription, accessibility, and voice interfaces.

**Image Generation:** Create images from text descriptions. DALL-E 3 offers the best prompt following, while Stable Diffusion provides more customization options. Consider content policies when enabling these.

**Code Models:** Specialized for programming tasks including code generation, explanation, debugging, and refactoring. Some are fine-tuned on specific languages or frameworks.

**Scientific Models:** Domain-specific models trained on scientific literature. Useful for research applications but require careful evaluation for accuracy.

### 6.3 Model Configuration

Click on a model to configure:

Setting	Description
<b>Enabled</b>	Available for use
<b>Min Tier</b>	Minimum subscription tier
<b>Rate Limits</b>	Requests per minute
<b>Max Tokens</b>	Maximum context/output
<b>Temperature Range</b>	Allowed temperature values
<b>Price Override</b>	Custom pricing

**Configuration Settings Explained Enabled:** When disabled, the model is hidden from users and API requests return “model not found”. Use this to temporarily remove models during maintenance or to restrict access to specific models.

**Min Tier:** Sets the minimum subscription tier required to access this model. For example, setting GPT-4 to Tier 2 means Free tier users cannot use it. This helps control costs and create upgrade incentives.

**Rate Limits:** Controls requests per minute per user for this model. Prevents abuse and ensures fair access. Set based on the provider’s rate limits and your capacity: - Conservative: 10-20 requests/minute - Standard: 50-100 requests/minute - High: 200+ requests/minute (requires provider rate limit increases)

**Max Tokens:** Limits context window and output length. Useful for controlling costs since longer contexts cost more. Set based on use case: - Short tasks (Q&A): 4,096 tokens - Medium tasks (writing): 16,384 tokens - Long tasks (analysis): 32,768+ tokens

**Temperature Range:** Restricts the temperature parameter users can set. Temperature controls randomness: - 0.0: Deterministic, consistent outputs - 0.7: Balanced creativity and consistency - 1.0+: More creative, less predictable

Restricting range (e.g., 0.0-1.0) prevents users from setting extreme values that produce poor results.

**Price Override:** Allows custom pricing different from the default. Useful for: - Offering discounts on specific models - Increasing prices for premium models - Matching competitor pricing - A/B testing pricing strategies

## 6.4 Self-Hosted Models

For Tier 3+ deployments:

1. Navigate to **Models** → **Self-Hosted**
2. Click “+ Add Self-Hosted Model”
3. Configure:
  - **Model ID:** Unique identifier
  - **SageMaker Endpoint:** Endpoint name
  - **Instance Type:** ml.g5.xlarge, etc.
  - **Auto-Scaling:** Min/max instances
4. Deploy model to SageMaker

## 6.5 Thermal States (Self-Hosted)

State	Description	Response Time
<b>HOT</b>	Always running	<100ms
<b>WARM</b>	Scaled down	<5s
<b>COLD</b>	Stopped	30-60s
<b>OFF</b>	Disabled	N/A

## 7. Provider Management

### 7.1 External Providers

Navigate to **Providers** to manage API integrations:

Provider	Models	Status	Health
<b>OpenAI</b>	12	Configured	99.9%
<b>Anthropic</b>	6	Configured	99.8%
<b>Google AI</b>	8	Configured	99.7%
<b>xAI</b>	2	Configured	99.5%

Provider	Models	Status	Health
DeepSeek	4	Not configured	-

## 7.2 Adding Provider Credentials

1. Click on provider name
2. Click “**Configure**”
3. Enter API credentials:
  - **API Key:** Provider API key
  - **Organization ID:** (if applicable)
  - **Base URL:** (for custom endpoints)
4. Click “**Test Connection**”
5. Click “**Save**”

## 7.3 Provider Health Monitoring

View real-time provider health:

Provider Health: OpenAI

Status: Healthy  
 Uptime (30d): 99.94%  
 Avg Latency: 245ms  
 P95 Latency: 520ms  
 Error Rate: 0.02%

Last 24 Hours:

12am          6am          12pm          6pm          12am

## 7.4 Fallback Configuration

Configure provider fallbacks:

1. Navigate to **Providers** → **Fallbacks**
2. Set priority order for each model category
3. Configure automatic failover rules
4. Set retry policies

# 8. Billing & Subscriptions

## 8.1 Subscription Tiers

Tier	Name	Monthly	Features
1	Free	\$0	Basic models, 1K requests
2	Starter	\$29	More models, 10K requests
3	Professional	\$99	All external models, 100K requests
4	Business	\$299	Priority support, 500K requests
5	Enterprise	\$999	Self-hosted, unlimited
6	Enterprise+	Custom	Custom SLAs, dedicated support
7	Ultimate	Custom	On-premise options

## 8.2 Credit System

Credits are the universal currency for AI usage:

Model Type	Cost per 1M Tokens
GPT-4o	500 credits
GPT-4o-mini	50 credits
Claude 3 Opus	600 credits
Claude 3 Sonnet	150 credits
Self-hosted	20 credits

## 8.3 Managing Subscriptions

Navigate to **Billing** → **Subscriptions**:

1. View current subscription
2. Upgrade/downgrade tier
3. Add credit packages
4. View invoices
5. Update payment method

## 8.4 Usage Reports

Generate usage reports:

1. Navigate to **Billing** → **Reports**
2. Select date range
3. Choose grouping (by tenant/model/user)
4. Export as CSV/PDF

## 8.5 Billing Alerts

Configure alerts for:

- Credit balance low
- Usage spike
- Approaching quota
- Failed payments

## 9. Storage Management

### 9.1 Storage Overview

Navigate to **Storage** to monitor:

#### Storage Overview

Total Used: 234.5 GB of 500 GB (47%)

#### By Type:

Documents: 120.3 GB (51%)  
Images: 45.2 GB (19%)  
Audio: 38.7 GB (17%)  
Video: 22.1 GB (9%)  
Other: 8.2 GB (4%)

#### Top Tenants:

1. Acme Corp 45.2 GB  
2. TechStart 32.1 GB  
3. DataCo 28.4 GB

### 9.2 Storage Tiers

Tier	Included	Additional
Free	1 GB	N/A
Starter	10 GB	\$0.10/GB
Professional	100 GB	\$0.08/GB
Business	500 GB	\$0.05/GB
Enterprise	2 TB	\$0.03/GB

### 9.3 File Management

Manage uploaded files:

- View file metadata
  - Download files
  - Delete files
  - Set retention policies
-

## 10. Orchestration & Neural Engine

### 10.1 Brain Router

The Brain Router automatically selects optimal models:

Factor	Weight	Description
<b>Cost</b>	30%	Price optimization
<b>Quality</b>	30%	Output quality
<b>Speed</b>	20%	Response latency
<b>Availability</b>	20%	Provider health

### 10.2 Neural Patterns

Configure orchestration patterns:

Pattern	Description	Use Case
<b>Single</b>	One model	Simple requests
<b>Fallback</b>	Primary + backup	High availability
<b>Parallel</b>	Multiple simultaneous	Consensus
<b>Chain</b>	Sequential models	Complex tasks

### 10.3 Workflow Templates

Create reusable workflows:

1. Navigate to **Orchestration** → **Workflows**
2. Click “+ **New Workflow**”
3. Define steps and conditions
4. Set triggers and parameters
5. Save and activate

---

## 11. Localization

### 11.1 Translation Management

Navigate to **Localization** to manage:

- Supported languages
- Translation strings
- AI translation settings

### 11.2 Supported Languages

Language	Code	Status
English	en	Default

Language	Code	Status
Spanish	es	Enabled
French	fr	Enabled
German	de	Enabled
Japanese	ja	Enabled
Chinese	zh	Enabled

### 11.3 AI Translation

Enable AI-powered translation:

1. Navigate to **Localization** → **Settings**
2. Enable “**AI Translation**”
3. Select translation model
4. Configure quality settings

## 12. Configuration Management

### 12.1 System Configuration

Navigate to **Configuration** to manage:

Category	Settings
<b>General</b>	Platform name, domain, timezone
<b>Email</b>	SMTP settings, templates
<b>Security</b>	Password policy, MFA settings
<b>API</b>	Rate limits, CORS settings
<b>Features</b>	Feature flags

### 12.2 Tenant Overrides

Allow tenant-specific configuration:

1. Navigate to **Configuration** → **Tenant Overrides**
2. Select tenant
3. Override specific settings
4. Save changes

### 12.3 SSM Parameters

System configuration is stored in AWS SSM:

Parameter	Description
/radiant/prod/database/url	Database connection
/radiant/prod/api/rate-limit	API rate limits
/radiant/prod/features/*	Feature flags

---

## 13. Security & Compliance

### 13.1 Security Dashboard

Navigate to **Security** to monitor:

Security Dashboard

Threat Level: Low

Active Threats: 0  
Failed Logins: 23 (last 24h)  
Suspicious IPs: 2 blocked  
MFA Adoption: 94%

Recent Alerts:

Unusual login location - user@acme.com (2h ago)  
Resolved: Brute force attempt blocked (5h ago)  
Resolved: API key rotated - tenant xyz (1d ago)

### 13.2 Anomaly Detection

Automatic detection of:

- Impossible travel (geographic anomalies)
- Session hijacking attempts
- Brute force attacks
- Unusual API patterns

### 13.3 Compliance Reports

Navigate to **Compliance** to generate:

Framework	Description
<b>SOC 2</b>	Service organization controls
<b>HIPAA</b>	Healthcare data protection
<b>GDPR</b>	EU data protection
<b>ISO 27001</b>	Information security

### 13.4 Generating Reports

1. Click “**Generate Report**”
2. Select framework

3. Choose date range
  4. Select metrics to include
  5. Generate PDF/CSV
- 

## 14. Cost Analytics

### 14.1 Cost Dashboard

Navigate to **Cost** to view:

Cost Analytics		Period: Last 30 Days
Total Spend:	\$12,456.78	(+12% vs last month)
Projected:	\$14,200.00	(this month)
By Provider:		
OpenAI:	\$6,234.56	(50%)
Anthropic:	\$3,456.78	(28%)
Self-hosted:	\$1,234.56	(10%)
Other:	\$1,530.88	(12%)
AI Recommendations:		
Switch 23% of GPT-4 calls to GPT-4-mini (save \$890/mo)		
Enable caching for repeated queries (save \$340/mo)		

### 14.2 Cost Alerts

Configure alerts:

- Daily budget exceeded
- Weekly spend spike
- Per-tenant limits
- Per-model thresholds

### 14.3 Cost Optimization

Review AI-powered recommendations:

1. Navigate to **Cost** → **Insights**
  2. Review suggestions
  3. Click “**Apply**” to implement (requires approval)
  4. Track savings over time
-

## 15. A/B Testing & Experiments

### 15.1 Experiment Dashboard

Navigate to **Experiments** to manage:

Experiment	Status	Variants	Sample Size
Model routing v2	Running	3	45,234
Prompt optimization	Running	2	12,456
Temperature test	Completed	4	89,123

### 15.2 Creating an Experiment

1. Click “+ New Experiment”
2. Configure:
  - **Name:** Descriptive name
  - **Hypothesis:** What you’re testing
  - **Variants:** Control + treatments
  - **Traffic Split:** Percentage per variant
  - **Success Metric:** What to measure
3. Set targeting rules
4. Start experiment

### 15.3 Statistical Analysis

View results with:

- Conversion rates per variant
  - Statistical significance (p-value)
  - Confidence intervals
  - Sample size recommendations
- 

## 16. Audit & Monitoring

### 16.1 Audit Logs

Navigate to **Audit** to view all actions:

Column	Description
<b>Timestamp</b>	When action occurred
<b>Actor</b>	Who performed action
<b>Action</b>	What was done
<b>Resource</b>	What was affected
<b>IP Address</b>	Source IP
<b>Details</b>	Additional context

## 16.2 Log Filtering

Filter by:

- Date range
- Actor (user/admin)
- Action type
- Resource type
- Severity level

## 16.3 Log Export

Export logs for compliance:

1. Set filter criteria
2. Click “**Export**”
3. Choose format (CSV/JSON)
4. Download file

## 16.4 Real-Time Monitoring

Navigate to **Monitoring** for:

- Live request stream
  - Error rate graphs
  - Latency percentiles
  - Active users count
- 

## 17. Database Migrations

### 17.1 Migration Workflow

RADIANT uses dual-admin approval for production migrations:

1. **Submit:** Admin submits migration
2. **Review:** Second admin reviews
3. **Approve:** Second admin approves
4. **Execute:** Migration runs
5. **Verify:** Automatic verification

### 17.2 Pending Migrations

Navigate to **Migrations** to see:

Database Migrations

Pending Approval:

#045 - Add user preferences table

Submitted by: `alice@company.com` (2 hours ago)  
[View SQL] [Approve] [Reject]

Recent Migrations:

- #044 - Cost tracking tables (applied 2024-12-24)
- #043 - Experiment framework (applied 2024-12-20)
- #042 - Security anomalies (applied 2024-12-15)

## 17.3 Approving Migrations

1. Review the SQL in “**View SQL**”
  2. Check for potential issues
  3. Click “**Approve**” or “**Reject**”
  4. Add comment explaining decision
- 

## 18. API Management

### 18.1 API Keys

Manage platform API keys:

1. Navigate to **Settings** → **API Keys**
2. View existing keys
3. Create new keys with scopes
4. Revoke compromised keys

### 18.2 Rate Limiting

Configure rate limits:

Level	Default	Configurable
<b>Global</b>	10,000/min	Yes
<b>Per-Tenant</b>	1,000/min	Yes
<b>Per-User</b>	100/min	Yes
<b>Per-Key</b>	60/min	Yes

### 18.3 Webhooks

Configure outgoing webhooks:

1. Navigate to **Settings** → **Webhooks**
2. Add webhook URL
3. Select events to send
4. Test webhook
5. Enable webhook

---

## 19. Troubleshooting

### 19.1 Common Issues

#### High Error Rate

1. Check **Providers** for unhealthy providers
2. Review **Audit** logs for patterns
3. Check **Monitoring** for load spikes
4. Verify API key validity

#### Slow Response Times

1. Check provider latency in **Providers**
2. Review model selection in **Orchestration**
3. Check for cold-start issues (self-hosted)
4. Verify database performance

#### Authentication Failures

1. Check user status in **Users**
2. Verify MFA configuration
3. Review **Audit** logs for login attempts
4. Check for IP blocks in **Security**

### 19.2 Support Resources

Resource	Description
<b>Documentation</b>	This guide + online docs
<b>Status Page</b>	status.radiant.example.com
<b>Support Email</b>	support@radiant.example.com
<b>Emergency</b>	+1-555-RADIANT

### 19.3 Log Locations

Service	Log Group
API Gateway	/aws/apigateway/radiant
Lambda	/aws/lambda/radiant-*
Admin Dashboard	/aws/cloudfront/admin
Database	/aws/rds/cluster/radiant

## Appendix: Quick Reference

### Keyboard Shortcuts

Shortcut	Action
G + D	Go to Dashboard
G + T	Go to Tenants
G + M	Go to Models
G + B	Go to Billing
G + A	Go to Audit
?	Show shortcuts

### Status Indicators

Icon	Meaning
	Healthy/Success
	Warning
	Error/Failed
	In Progress
	Disabled/Pending

---

*Document Version: 4.18.1 Last Updated: December 2024*