

Accelerating AI-Enabled Research with Secure AI Workspaces at a Leading Research University

Customer Overview

The University is a leading research institution whose faculty and researchers increasingly rely on Large Language Models (LLMs) for data analysis, software development, scientific computing, and AI-assisted research.

Researchers wanted the flexibility to experiment with both open-source foundation models and commercial models without the complexity of manually configuring GPU infrastructure or managing AI software environments.

The University also wanted to ensure that commercial AI services were consumed securely through Amazon Bedrock while maintaining institutional governance, project isolation, and visibility into AI usage costs.

Relevance Lab partnered with the University to deploy AI Workspaces within Research Gateway, providing researchers with secure, self-service AI environments capable of running both local open-source models and Amazon Bedrock-hosted foundation models from a single integrated workspace.

Business Challenge

The University wanted to democratize access to AI technologies while simplifying infrastructure management and maintaining institutional governance.

Several key challenges needed to be addressed.

Challenge 1: Supporting Multiple AI Model Types

Challenge 2: Simplifying AI Infrastructure

Challenge 3: Flexible GPU Computing

Challenge 4: Secure Enterprise AI Access

Challenge 5: AI Cost Visibility

At a Glance

Industry

Higher Education & Research

Solution

Secure AI Research Workspace Platform

Platform

Relevance Lab Research Gateway on AWS

Primary Goals

AI Research Enablement, Secure Model Access, GPU Computing, Cost Governance

Research Areas

Data Science, Biomedical Research, Engineering, Social Sciences, AI & Machine Learning

Key Capabilities

AI Workspaces, Open-Source LLMs, Amazon Bedrock Integration, GPU Workstations, Self-Service Provisioning

Compliance Focus

Institutional Governance, Secure AI Access, Cloud Cost Visibility



Challenge 1

Supporting Multiple AI Model Types

Different research projects required different types of AI models.

Some researchers wanted to execute open-source models locally on GPU-enabled servers using Ollama, while others preferred managed commercial models such as Claude or GPT through Amazon Bedrock.

The institution wanted a single research platform capable of supporting both approaches.

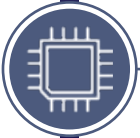


Challenge 2

Simplifying AI Infrastructure

Installing GPU drivers, AI frameworks, model runtimes, and development environments required significant technical expertise.

Research IT wanted researchers to begin using AI immediately without spending hours configuring software or infrastructure.



Challenge 3

Flexible GPU Computing

Researchers required the ability to provision GPU-enabled infrastructure only when needed and to select the most appropriate GPU configuration based on their workload.

This flexibility would allow the university to optimize cloud costs while providing access to high-performance computing resources.



Challenge 4

Secure Enterprise AI Access

Commercial AI services needed to be consumed securely through Amazon Bedrock, allowing institutional policies to determine which foundation models were available to researchers.

All Amazon Bedrock API calls needed to remain within the university's AWS environment while maintaining centralized governance.



Challenge 4

AI Cost Visibility

Research administrators required visibility into both infrastructure costs and Amazon Bedrock API consumption.

AI-related expenses needed to be attributed to individual research workspaces and included as part of overall project cost reporting.

The objective was to provide researchers with a modern AI research environment that combined open-source innovation with enterprise AI services while minimizing operational complexity for Research IT.



AI Workspaces on AWS

Relevance Lab deployed Research Gateway with a new AI Workspaces catalog offering that enabled researchers to provision fully configured AI development environments through a self-service portal.

Each workspace combined GPU computing, AI development tools, open-source LLM support, and Amazon Bedrock integration within a standardized research environment.

Self-Service AI Workspace Provisioning

Researchers could provision AI-enabled workstations on demand using Research Gateway.

Key capabilities included:

- Browser-based graphical remote desktop
- Choice of CPU or GPU-enabled EC2 instances
- NVIDIA GPU drivers pre-installed
- Standardized software images
- Persistent research storage
- Self-service lifecycle management

The entire environment was automatically configured during provisioning, allowing researchers to begin productive work immediately.

Open-Source AI Models with Ollama

Every AI Workspace included Ollama deployed as a Docker container, providing a simple runtime for executing open-source large language models locally.

Researchers could:

- Access Ollama through a local API endpoint
- Run inference entirely within their workspace
- Experiment with multiple open-source models
- Perform private AI inference without external dependencies

The desired open-source model could be specified during workspace provisioning and was automatically downloaded and configured during setup.

Integrated AI Development Environment

Each workspace included a comprehensive collection of research tools.

Pre-installed software included:

- Python
- R
- JupyterLab
- RStudio
- Git
- Ollama
- OpenWebUI

OpenWebUI was deployed as a Docker container, providing an intuitive browser-based chat interface for interacting with locally hosted AI models without requiring command-line expertise.

Amazon Bedrock Integration

In addition to locally hosted models, every workspace was configured with secure access to Amazon Bedrock.

Instance IAM roles enabled access only to foundation models approved by Research IT.

This approach allowed researchers to seamlessly choose between:

- Local open-source models running through Ollama
- Managed foundation models hosted on Amazon Bedrock

All Amazon Bedrock API requests remained entirely within the university's AWS environment.

Research Data Integration

Research Gateway automatically mounted project study folders into every AI Workspace.

Researchers could securely access approved research datasets while benefiting from the platform's existing governance, permissions, and project isolation capabilities.

Integrated AI Cost Monitoring

Research Gateway extended its financial reporting capabilities to include Amazon Bedrock API consumption.

The platform automatically tracked:

- Infrastructure costs
- GPU compute costs
- Amazon Bedrock API costs
- Workspace operating costs
- Project-level AI spending

This provided administrators with complete visibility into the total cost of AI-enabled research.

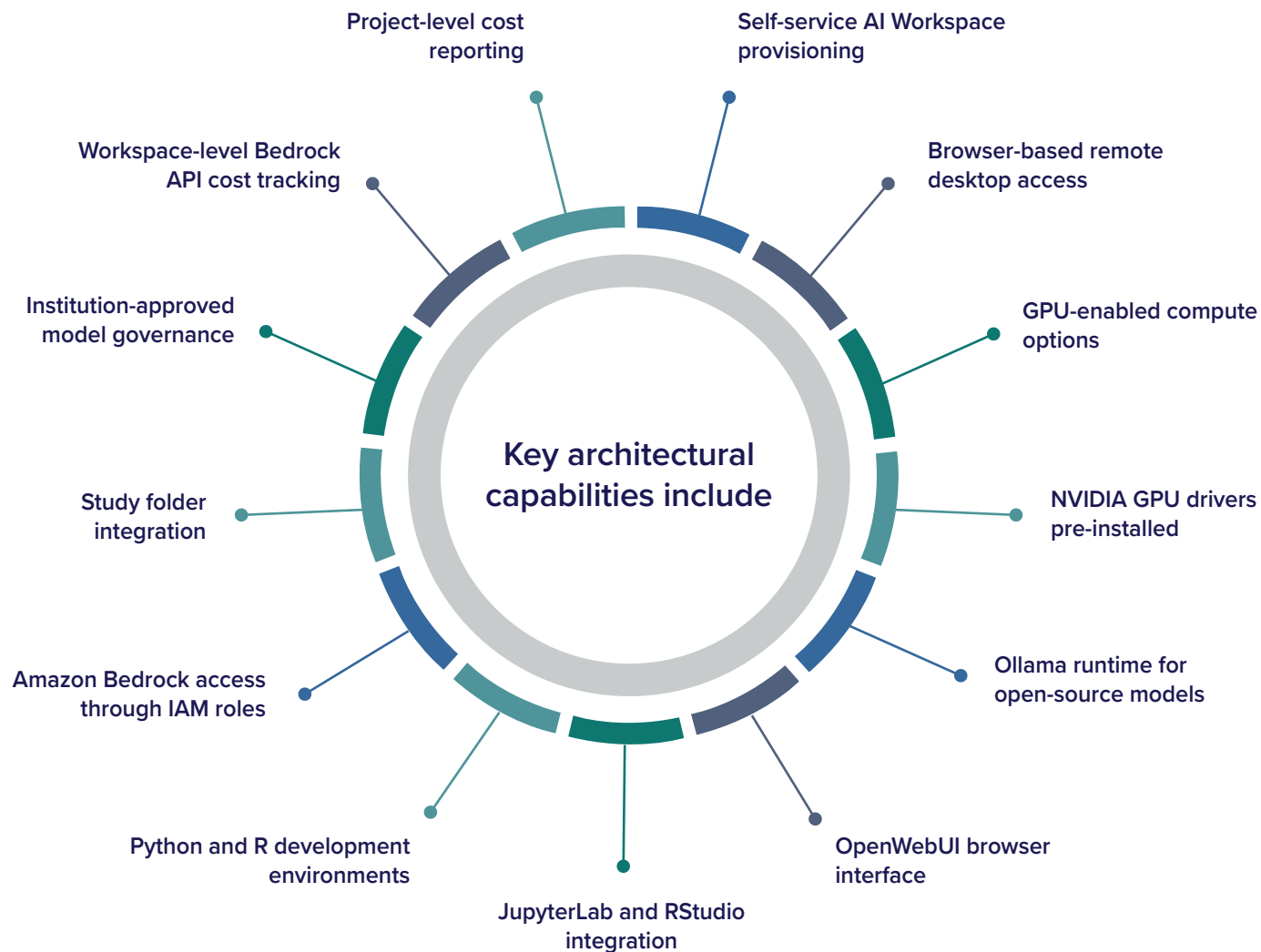


Architecture Highlights

Unified AI Research Platform

Researchers authenticate through Research Gateway and provision AI Workspaces using institution-approved catalog offerings.

Each workspace combines GPU infrastructure, local AI model execution, enterprise AI services, and research collaboration tools within a secure cloud environment.



Business Outcomes

Deployment of AI Workspaces enabled the University to rapidly introduce AI-powered research capabilities while simplifying infrastructure management and maintaining institutional governance.

Researcher Productivity

Researchers gained immediate access to fully configured AI development environments capable of supporting a broad range of AI workflows.

Benefits included:

- Rapid provisioning of AI-ready workspaces
- Immediate access to GPU-enabled computing
- Simplified experimentation with multiple foundation models
- Reduced environment configuration time
- Faster onboarding for AI research projects

Flexible AI Innovation

Researchers could choose the AI technology best suited to their work without requiring separate platforms.

Capabilities included:

- Local execution of open-source models using Ollama
- Browser-based interaction through OpenWebUI
- Enterprise AI services through Amazon Bedrock
- Integrated notebooks and statistical computing environments
- Secure access to institutional research data

Operational Efficiency

Research Gateway automated deployment and lifecycle management of AI Workspaces, significantly reducing administrative effort.

Administrative improvements included:

- Self-service provisioning
- Standardized AI workstation images
- Automated software installation
- Centralized governance of Amazon Bedrock models
- Reduced manual infrastructure support

Financial Visibility

Research Gateway provided comprehensive reporting across both infrastructure and AI consumption.

This enabled the University to:

- Monitor Amazon Bedrock API usage by workspace
- Allocate AI costs to research projects
- Track GPU utilization
- Improve cloud budgeting
- Optimize AI infrastructure spending

Customer Perspective

“

Research Gateway enabled the University to rapidly deliver enterprise-grade AI capabilities to our researchers without introducing unnecessary operational complexity. By combining open-source models, Amazon Bedrock, GPU computing, and self-service provisioning into a single platform, we have accelerated AI adoption while maintaining governance, cost visibility, and institutional security.”

— University Research Computing Leadership



Key Outcomes

- ✓ Rapid deployment of AI-enabled research environments
- ✓ Self-service provisioning of GPU-enabled AI Workspaces
- ✓ Support for both open-source models through Ollama and enterprise models through Amazon Bedrock
- ✓ Browser-based OpenWebUI interface for conversational AI
- ✓ Integrated JupyterLab and RStudio environments
- ✓ Automated deployment of researcher-selected AI models
- ✓ Secure access to Amazon Bedrock with institution-controlled model governance
- ✓ Research Gateway Study Folder integration for secure data access
- ✓ Per-workspace Amazon Bedrock API cost tracking
- ✓ Unified reporting of infrastructure and AI operating costs
- ✓ Reduced administrative effort and accelerated AI adoption

About Relevance Lab Research Gateway

Secure AI. Flexible Infrastructure. Built for Research Innovation.

Research Gateway is Relevance Lab's cloud-native platform for secure research computing, AI-enabled workspaces, and scientific collaboration. The platform enables universities, healthcare organizations, government agencies, and research institutions to rapidly deploy governed cloud workspaces, integrate enterprise AI services, support open-source AI frameworks, optimize cloud and AI costs, and accelerate research while maintaining institutional security and compliance.

[Learn More About Research Gateway](#)

Recognized as a
trusted AWS partner

