



Case Study

How did WeCredit and AWS build a private WhatsApp GenAI platform with continuous fine-tuning?

Solution At a Glance

WeCredit, a leading digital lending and credit-risk automation company, partnered with AWS to build a **secure, scalable and fully private AI-driven conversational automation platform** for borrower onboarding and real-time support. By leveraging a custom VPC architecture, Amazon Bedrock, OpenSearch, Lambda, DynamoDB and EventBridge, WeCredit now delivers **low-latency AI responses**, secure data processing, and automated model retraining while ensuring governance, auditability and enterprise-level reliability.

Key Outcomes Delivered

- **Highly secure AI automation** with private VPC networking and KMS encryption.
- **Real-time inference pipeline** with Bedrock, Lambda and API Gateway.

- **Accurate and compliant responses** powered by Knowledge Bases and prompt governance.
- **Automated data processing & model refinement** via Glue, EventBridge & Bedrock fine-tuning.
- **Human-in-the-loop workflow** for quality control using Amazon A2I & SQS.
- **Resilient, monitored system** with centralized CloudWatch dashboards and alerts.
- **Well-Architected foundation** following AWS best practices across security, reliability and cost.

About Client and their Mission

WeCredit is a digital lending and financial automation platform that helps NBFCs, financial institutions and fintechs manage borrower queries, risk analysis and credit workflows.

With rapid user growth, rising loan volumes and evolving compliance requirements, WeCredit needed an **AI-powered automation system** that stays **secure, fast, accurate and fully governed**.

Core Challenges Addressed

As WeCredit scaled, the organization faced several critical hurdles:

- Ensuring **zero public exposure** of AI models, data pipelines or storage layers.
- Maintaining **encrypted, auditable processing** for regulated financial data.
- Deploying **low-latency AI inference** inside a private multi-AZ architecture.
- Managing **high-volume conversational traffic** for customer queries.
- Building a **continuous data refresh and model update** workflow.
- Implementing **automated governance** for prompts, training and knowledge bases.

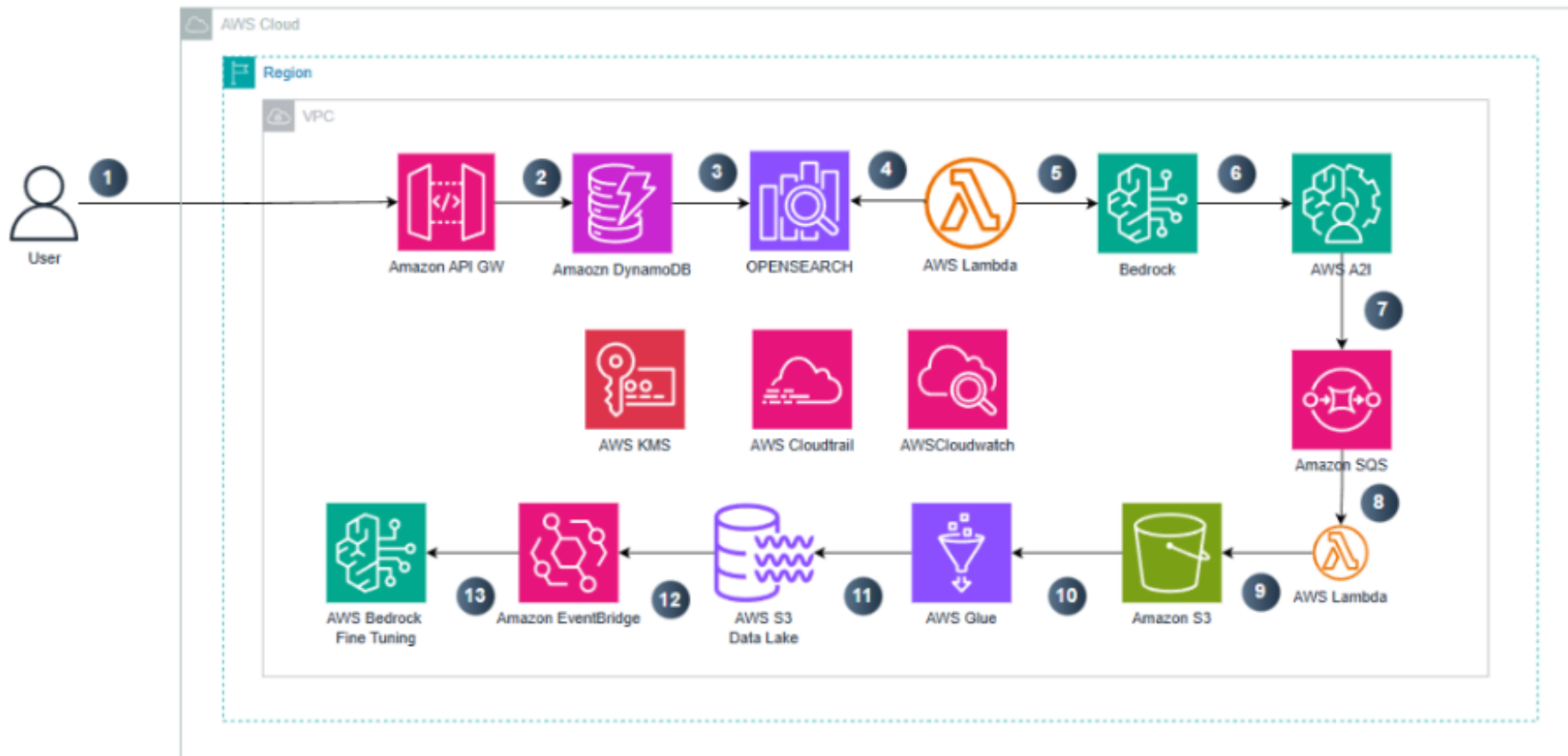
- Enabling **human review** where required for compliance-sensitive responses.

WeCredit requires a **fully private, production-grade Generative AI platform** aligned with AWS Responsible AI best practices and the Well-Architected Framework.

Proposed AWS Solutions

- Implemented a **fully private, multi-AZ AWS architecture** for secure, reliable and real-time GenAI automation in financial workflows.
- Built a **custom VPC** with private subnets, NAT Gateways and restrictive routing to ensure **no AI or data services are publicly exposed**.
- Routed requests via **API Gateway** to a **VPC-enabled Lambda** that orchestrates inference by managing session context, validating inputs and calling **Amazon Bedrock**.
- Stored conversation state in **DynamoDB** and enabled semantic retrieval using **OpenSearch** for domain documents and embeddings.
- Created an **S3 data lake** for raw, processed and training datasets with **KMS encryption** and **Glue ETL** for periodic data preparation.
- Automated **model retraining and knowledge refresh with EventBridge** using Bedrock fine-tuning workflows to stay aligned with evolving rules and compliance needs.
- Added **Human in the Loop review** using **Amazon A2I** and **SQS** for flagged responses before delivery to end users.
- Ensured governance and observability with **CloudWatch** dashboards, logs and alerts plus **CloudTrail** for auditability backed by **least-privilege IAM** and end-to-end **KMS encryption**.

End-to-End Architectural Diagram



Result and Impact

- **Enhanced Security & Compliance**

- Deployment inside a **fully private VPC** ensures no public exposure.
- **KMS encryption** applied across S3, DynamoDB, OpenSearch and logs.
- Comprehensive traceability with **CloudTrail** and governance workflows.

- **Real-Time AI Response System**

- Bedrock + Lambda pipeline ensures **fast conversational responses**.
- DynamoDB + OpenSearch provide **context-rich, accurate answers**.

- **Scalable, Fault-Tolerant Architecture**

- Multi-AZ infrastructure ensures **high availability**.
- Serverless components automatically scale to demand.

- **Continuous Model Improvement**
 - Automated data refreshed via EventBridge schedules.
 - Glue ETL jobs prepare datasets for **Bedrock fine-tuning**.
 - Knowledge Bases update automatically based on S3 changes.
- **Human-in-the-Loop Assurance**
 - A2I-based workflow ensures **review where accuracy & compliance are critical**.
- **Operational Excellence & Monitoring**
 - CloudWatch dashboards monitor **latency, errors, throttles, queue depth, model performance**, etc.
 - Comprehensive alerting ensures proactive mitigation.
- **Cost-Efficient & AWS Well-Architected**
 - Serverless-first approach reduces overprovisioning.
 - Adheres to all six Well-Architected pillars.

About Us



Years of Exp

15+



Data Center

10+



Awards

100+

AceCloud is a leading provider of end-to-end cloud computing solutions to global organizations across industries at scale. It offers a full spectrum of cloud services, including Public Cloud, Application Hosting, AWS Services, Managed Security Services, and Hosted Virtual Desktop Solutions.

AceCloud is a brand of Real Time Data Services (RTDS) group of companies – a leading provider of information technology capabilities specializing in Cloud Computing and Cloud Telephony. RTDS Group has an employee base of 600+ across India, the US, and the UK, supporting over 20,000+ customers and IT infrastructure in 10+ data centers spanning the globe. It offers industry-leading technological solutions that help customers streamline their operations and enhance efficiency.

As an advanced consulting and certified AWS Well-Architected Partner, AceCloud recently achieved the prestigious AWS Storage Competency and is also recognized as an Amazon RDS Service Delivery Partner. These achievements underscore its deep expertise in AWS technologies and commitment to delivering high-quality cloud solutions.



- Public Sector
- Storage Consulting Competency
- Small and Medium Business Services Competency
- Amazon RDS Delivery
- AWS Database Migration Service Delivery
- Amazon CloudFront Delivery
- AWS Well-Architected Partner Program
- Authorized Commercial Reseller



Contact Us

Are you ready to cut down your cloud spending? Contact AceCloud Experts today to get a customized quote.



awspartner@acecloud.ai



<https://acecloud.ai/aws/>

Our Location

2069 Oneill Dr, Bethel
Park, Pennsylvania,
15102-6602

2637 E Atlantic Blvd
#23519 Pompano Beach,
FL 33062, Florida

Ace Tower, 809-A, Udyog
Vihar, Phase 5, Gurugram
122016, Haryana, India