source allies

# Measurable AI-Driven Efficiency

## SUMMARY

Our client, an established roofing and building envelope solutions company, sought to reduce the workload of their customer support & sales teams, who are tasked with answering complex, technical questions across their extensive product line. These inquiries often involve detailed data specifications, precise measurements, and evolving application or safety protocols.

## SOLUTION

We built a custom Retrieval Augmented Generation (RAG) solution using an Agentic AI framework to aggregate 800+ documents and deliver comprehensive answers to complex queries. We integrated custom internal documentation, preserving institutional knowledge, while providing rapid, accurate responses via a user-friendly interface for specialists. For data security, we implemented AWS Cognito with Google as the Identity Provider, ensuring protection from data ingestion through user interactions.

To improve reliability, we collaborated with the client's Subject Matter Experts to validate FAQs and factual answers, incorporating them into our Metrics-Driven Development approach. This allowed real-time tracking of system performance, ensuring ongoing accuracy. Our custom Gen AI solution went live in 5 weeks when the client previously could not get the OpenAI based solution into production at all. As a result, the client trusted the AI solution's ability to reduce manual workload while maintaining confidence in its outputs.

**Tech Stack:** Tech Stack: Amazon API Gateway, AWS Lambda, Amazon SQS, Amazon S3, Amazon DynamoDB, Amazon Bedrock, Claude v3 Sonnet model for 'chat', Titan v1 model for embeddings, Amazon RDS (Postgres), AwS Cognito.

## RESULTS

**5 weeks**

Start date to POC meeting client benchmarks

**85%**

Hitting 85% answer correctness in POC by week 5

**94.9%**

Hitting 94.9% answer similarity in POC