TRIANZ℠

Project Story

Agentic AI Based Policy Agent

Trianz_AWS Hackathon '25

trianz.com

TABLE OF CONTENTS

Inspiration

The healthcare insurance industry faces a critical challenge: complex policy documents that are difficult for customers to understand, leading to poor decision-making and increased claim disputes. Traditional insurance advisors are expensive and not available 24/7. We envisioned an intelligent, conversational AI agent that could democratize access to insurance expertise, making policy selection as simple as talking to a trusted friend.

Our inspiration came from observing how people struggle with insurance jargon, premium calculations, and policy comparisons. We asked ourselves: "What if we could create an AI agent that not only reads and understands complex insurance documents but also explains them in natural language, provides personalized recommendations, and ensures regulatory compliance—all through voice interaction?"

This vision led us to build Alan, a sophisticated AI-powered health insurance advisor that combines AWS Bedrock's Amazon Nova models with advanced agentic workflows to deliver intelligent, explainable, and compliant insurance guidance.

What It Does

This project is an enterprise-grade, multi-agent AI system that revolutionizes health insurance policy management and customer advisory services. The system provides:

Conversational Policy Advisory

Users interact with Alan through natural voice conversations powered by Amazon Nova Sonic, discussing their insurance needs, asking questions about policy details, and receiving personalized recommendations in real-time.

Intelligent Document Processing

The system automatically ingests, validates, and analyses complex health insurance policy documents using a coordinated team of eight specialized AI agents, each focusing on specific aspects like medical risk assessment, financial analysis, compliance verification, and lifestyle factors.

Explainable AI Decision-Making

Every recommendation comes with transparent reasoning, showing users exactly how premiums were calculated, why specific coverage was recommended, and what factors influenced the decision—building trust through transparency.

Automated Policy Generation

Upon successful underwriting analysis, the system automatically generates professionally formatted PDF policy documents with all necessary legal clauses, coverage details, and personalized terms.

Regulatory Compliance

Built-in compliance agents ensure all processing adheres to HIPAA regulations, state insurance laws, and company policies, with automatic PHI masking and audit trail generation.

Real-Time Status Tracking

A sophisticated S3-based status tracking system provides live updates on document processing, allowing frontend applications to display progress as each agent completes its analysis.

How We Built It

Architecture Foundation

We architected SDIS as a hybrid cloud-native application leveraging AWS services extensively. The frontend runs on Flask with SocketIO for real-time bidirectional communication, while the backend processing occurs entirely on AWS Bedrock AgentCore.

Multi-Agent Orchestration

We designed eight specialized Nova Pro agents, each with carefully crafted system prompts encoding domain expertise:

1. **Data Intake Agent**: Handles document ingestion and initial categorization
2. **Document Verification Agent**: Validates authenticity and completeness
3. **Medical Risk Assessment Agent**: Evaluates health-related mortality risks
4. **Financial Analysis Agent**: Assesses coverage appropriateness and affordability
5. **Driving Record Agent**: Analyzes transportation-related risk factors
6. **Compliance Verification Agent**: Ensures regulatory adherence
7. **Lifestyle Behavioural Analysis Agent**: Evaluates behavioural risk patterns
8. **Summary Generation Agent**: Produces comprehensive HTML reports

The orchestrator agent coordinates this workflow, ensuring all agents execute sequentially and their insights are properly synthesized.

- **Voice Integration**: We integrated Amazon Nova Sonic for natural voice interactions, implementing a sophisticated conversation handler that manages bidirectional audio streaming, speech-to-text conversion, and contextual conversation tracking. The system intelligently extracts policy preferences, personal information, and health details from natural conversations.
- **Document Processing Pipeline**: We built a robust PDF extraction system using PyPDF2 that handles various document formats, encoding issues, and corrupted files gracefully. Documents are processed page-by-page with error recovery mechanisms.

- **Session Management**: We implemented dynamic session creation with unique timestamped identifiers, storing all session data and conversation context in S3. This enables persistent tracking across the entire customer journey—from initial conversation through document upload to final policy generation.
- **Policy Generation Engine**: We developed an automated policy generation system that extracts structured data from underwriting summaries using Nova Pro, then generates professional PDF documents with ReportLab, complete with tables, styling, and legal formatting.
- **Real-Time Updates**: We built an S3-based status tracking mechanism where each agent writes its completion status to a JSON file. The Flask frontend monitors this file through continuous polling, updating the UI in real-time as agents complete their work.
- **Deployment Architecture**: The entire backend is deployed on AWS Bedrock AgentCore, which provides serverless execution, automatic scaling, and built-in security. We packaged the application with all dependencies and configuration in the AgentCore format for seamless deployment.

Challenges We Ran Into

Credential Management Complexity

Managing AWS temporary session tokens across multiple boto3 clients proved extremely challenging. We had to implement comprehensive credential validation, automatic refresh mechanisms, and graceful degradation when credentials expire. The STS token lifecycle required careful handling to prevent mid-processing failures.

Agent Coordination and Error Recovery

Ensuring all eight agents execute reliably in sequence, even when individual agents encounter errors, required sophisticated error handling. We implemented a "never stop" philosophy where the orchestrator continues processing regardless of warnings, consolidating all findings in the final summary.

Nova Sonic Audio Streaming

Implementing real-time bidirectional audio streaming with Nova Sonic was technically complex. We had to manage audio chunk encoding, handle WebSocket connections, coordinate speech recognition timing, and ensure smooth conversation flow without interruptions or overlaps.

Encoding Issues in PDF Extraction

Health insurance documents often contain special characters, non-standard fonts, and various encodings. We encountered numerous UnicodeDecodeError exceptions that required implementing robust encoding handling with fallback mechanisms and graceful error messages.

Context Preservation Across Agents

Maintaining conversation context and document content across eight sequential agent calls while avoiding token limit issues required careful prompt engineering. We had to balance providing sufficient context versus staying within model token limits.

S3 Status Synchronization

Implementing real-time status updates through S3 required careful consideration of eventual consistency, race conditions, and polling frequency. We had to ensure the frontend accurately reflects backend processing state without overwhelming S3 with requests.

Policy Generation Accuracy

Extracting structured policy data from free-form HTML summaries required multiple prompt iterations. We needed Nova Pro to consistently produce valid JSON with specific fields, requiring detailed extraction rules and validation logic.

Voice Conversation Flow Control

Preventing the AI from interrupting users or providing upload links prematurely required sophisticated conversation state management. We implemented explicit trigger detection that only responds to clear user requests rather than automatic assumptions.

Accomplishments That We're Proud Of

Production-Ready Multi-Agent System

We successfully built and deployed a fully functional eight-agent underwriting system that processes real insurance documents end-to-end, from intake through policy generation, with enterprise-grade reliability.

Seamless Voice Experience

The Nova Sonic integration delivers remarkably natural conversations where users can speak freely about their insurance needs, and Alan responds intelligently with appropriate follow-up questions and helpful information.

True Explainable AI

Every underwriting decision comes with transparent, detailed reasoning. Users can see exactly which factors influenced their premium, what medical conditions were considered, and how their financial situation impacted coverage recommendations.

Automated Compliance

The system automatically ensures HIPAA compliance, performs PHI detection and masking, validates regulatory requirements, and maintains complete audit trails—critical for insurance industry deployment.

Professional Policy Documents

The automated policy generation produces insurance-grade PDF documents that meet legal standards, complete with proper formatting, tables, legal clauses, and all required disclosures.

Hybrid Architecture Excellence

We achieved seamless integration between local Flask frontend and cloud AgentCore backend, with S3 serving as the coordination layer, demonstrating sophisticated distributed systems design.

Robust Error Handling

The system gracefully handles credential expiration, model failures, document corruption, encoding issues, and network problems, always providing helpful error messages and recovery guidance.

Real-Time Status Visibility

The live agent status tracking gives users transparency into the processing pipeline, showing exactly which agents are working and what they're analysing at any moment.

What We Learned

Agent Design Patterns

We learned that successful multi-agent systems require careful responsibility separation, with each agent having a clearly defined scope. Overly broad agent responsibilities lead to inconsistent outputs, while properly scoped agents produce focused, high-quality results.

Prompt Engineering Mastery

Effective system prompts must encode not just domain knowledge but also output format requirements, business rules, and decision-making frameworks. We learned to be extremely specific about expected behaviours, especially regarding when to continue versus stop processing.

AWS Credential Management

Working extensively with temporary credentials taught us the importance of comprehensive validation, proper error messaging, and graceful degradation. We learned to always check credential validity before expensive operations.

Conversational AI Nuances

Building natural voice experiences requires deep understanding of conversation state management, turn-taking, interruption handling, and context tracking. We learned that explicit user intent detection is more reliable than assumptions.

Document Processing Resilience

Production document processing must handle every edge case: corrupted PDFs, unusual encodings, missing pages, scanned images as PDFs, and password-protected files. Robust error recovery is not optional—it's essential.

S3 as Coordination Layer

We learned that S3 can effectively serve as a state management layer for distributed systems, enabling coordination between frontend and backend with proper JSON structure and consistent update patterns.

Importance of Observability

Comprehensive logging at every stage—agent execution, S3 updates, document processing, voice interactions—proved invaluable for debugging and understanding system behaviour in production scenarios.

Model Selection Matters

Different Nova models serve different purposes. Nova Pro excels at complex reasoning and structured output, Nova Sonic provides superior voice experiences, and understanding these strengths allows optimal model selection per task.

What's Next for Agentic AI Based Policy Agent

Enhanced Personalization

We plan to implement customer profile learning where Alan remembers past conversations, preferences, and family situations across sessions, providing increasingly personalized recommendations over time.

Multi-Language Support

Expanding to Spanish, Mandarin, and other languages to serve diverse communities, with culturally appropriate insurance explanations and terminology.

Integration with Insurance Carriers

Direct API connections to major insurance carriers for real-time premium quotes, policy comparison, and instant enrolment, transforming Alan from an advisor to a complete insurance marketplace.

Advanced Risk Prediction

Implementing predictive models that forecast future insurance needs based on life stage, health trends, and family planning, proactively recommending coverage adjustments.

Mobile Application

Native iOS and Android apps with offline voice capabilities, document scanning, and push notifications for policy updates and renewal reminders.

Claims Assistance

Expanding Alan's capabilities to guide users through the claims process, helping gather required documentation, filling out forms, and tracking claim status.

Telemedicine Integration

Connecting with telehealth platforms to schedule appointments, coordinate care, and ensure medical records are properly submitted for underwriting.

Continuous Learning Pipeline

Implementing feedback loops where underwriters review Alan's decisions, with corrections automatically improving the model through fine-tuning and RAG updates.

Blockchain-based Policy Records

Exploring distributed ledger technology for immutable policy records, ensuring transparency and preventing fraud while maintaining privacy.

Thank You

reach@trianz.com