AI-Powered Talent Database with Agentic Chat for IO Progression

Executive Summary

IO Progression is a professional development and career services organization that manages a growing talent pool of certified professionals. Nova built a complete talent management platform for the customer, and at its core, an agentic chat interface that enables recruiters to interact with the system through natural language.

Nova implemented the agentic chat using **Amazon Bedrock** and **Strands Agents SDK**. The orchestrator agent interprets intent, selects appropriate tools, and executes a single tool per conversation turn through guided, multi-turn interactions.

The agentic solution delivers:

- Natural language interaction replacing manual searches and form-based workflows.
- Autonomous action execution with human confirmation on critical decisions.
- Intelligent candidate discovery with explainable reasoning based on stored resume data.
- Contextual multi-turn conversations with recruiter oversight.

The Challenge

IO Progression's recruiters managed talent data through scattered OneDrive resumes and manual Excel hiring matrices. The core problem was not data storage but interaction complexity: recruiters needed to discover candidates by combining multiple criteria, understand why a candidate matches, execute actions (create positions, start processes) from conversation, and ask contextual follow-up questions. Traditional search interfaces and form-based applications could not deliver this experience.

Why AWS?

AWS was selected for its fully managed, serverless AI services with pay-per-use pricing and enterprise-grade security for candidate PII. Amazon Bedrock provided access to multiple foundation models from a single managed service, and Strands Agents SDK, an open-source framework from AWS, enabled the orchestrator-with-tools pattern optimized for Lambda execution.

About Costumer

IO PROGRESSION
SPREADING KNOWLEDGE

Figure 1 – IO Progression Logo

IO Progression is a professional development and career services organization. Their mission is to help professionals find the right career path through professional certificates, hands-on projects, and skills recognition programs that achieve job-readiness in in-demand career fields.

"Skill up, prove your expertise to employers and peers, and get the recognition and opportunities you've earned.".

"Amazon Bedrock"

By leveraging Amazon Bedrock's foundation models and Strands Agents SDK, Nova was able to focus on building an intelligent conversational experience for recruiters while ensuring reliable, explainable AI recommendations grounded exclusively in stored candidate data.

Figure 2 – Amazon Bedrock Service

The Solution

Orchestrator Agent Architecture

The core of the solution is an orchestrator agent powered by **Claude 3.5 Haiku** through the **Strands Agents SDK**, an open-source framework from AWS. Claude was selected over other Bedrock models for its mature native tool_use capability, which enables single-call tool selection without additional parsing or adapter layers, and its ability to handle increasing complexity as the number of tools scales.

The orchestrator analyzes the recruiter's query, selects the correct tool, executes it, and responds. Each turn handles a single tool execution. For queries requiring extended processing, an async job pattern maintains responsiveness: Lambda returns a jobId immediately and the client polls for results.

Specialized Tool Agents

The orchestrator routes to three domain-specific tool agents:

- **RAG Candidate Tool Agent**: semantic candidate search using Amazon Titan Embeddings v2 and Amazon OpenSearch Serverless. For complex ranking, invokes Claude 3.5 Sonnet to provide explained recommendations.
- **Position Tool Agent**: position creation and retrieval from conversation, validating completeness and confirming before execution.
- **Hiring Process Tool Agent**: links candidates to positions, resolving references from conversation context and confirming with the recruiter.

Each tool agent maintains its own service clients, ensuring domain isolation. This encapsulation enables future evolution: any tool agent can be promoted to a full autonomous agent with its own LLM without modifying the orchestrator.

Conversational Workflow Design

The system is multi-turn conversational, not multi-step autonomous. Each turn executes a single tool, and the orchestrator prioritizes safety and confirmation over full autonomy:

- Validates that all required fields are present before executing actions.
- If information is missing, asks for it in a single message.
- Confirms all details with the recruiter before irreversible operations.
- Maintains conversation context across turns, enabling natural follow-up references.
- Provides explainable reasoning for candidate recommendations rather than opaque scores.

Responsible AI Controls

- Human-in-the-loop: confirmation required before state-changing operations (create position, start hiring process).
- Prompt injection protection: fixed tool scope, no dynamic code execution, typed parameters.
- AI output validation: candidate data always fetched from DynamoDB, never generated by the LLM.
- Session isolation: unique session IDs per chat, no cross-user data leakage.
- Query manipulation prevention: orchestrator uses exact search terms without query expansion.
- Cost and token protection: max_tokens limits, Bedrock usage alerts, execution timeouts.

Best Features

- **Conversational Workflow Execution**
- **Autonomous Tool Selection**
- **Explainable AI Reasoning**
- **Human-in-the-Loop Controls**
- **Contextual Conversation**
- **Extensible Tool Agent Pattern**

Figure 3 – Strands Agents SDK

Results and Benefits

Amazon Bedrock's Claude 3.5 models were central to the success of this agentic solution. Claude's native tool_use capability enables the orchestrator to analyze intent and select the appropriate tool in a single LLM call, achieving the single-turn execution pattern critical for serverless constraints. This capability is not available in other Bedrock models (Llama, Mistral, Titan Text) at the maturity level required for production recruitment workflows.

The agentic chat interface has fundamentally changed how recruiters interact with talent data. Instead of constructing search filters, navigating forms, and manually chaining operations, recruiters now describe what they need in natural language and the orchestrator autonomously executes the workflow. This shift from manual navigation to conversational delegation reduces cognitive load and accelerates time-to-action.

Explainability builds trust in AI recommendations

When the RAG agent returns candidates, it explains why each matches (specific skills, experience level, role alignment) rather than presenting opaque similarity scores. These explanations are based exclusively on data extracted from resumes stored in the system; the LLM never fabricates candidate information, it only filters and ranks existing data. Recruiters understand the reasoning and can make informed decisions about which candidates to pursue.

Human-in-the-loop controls ensure recruiter oversight

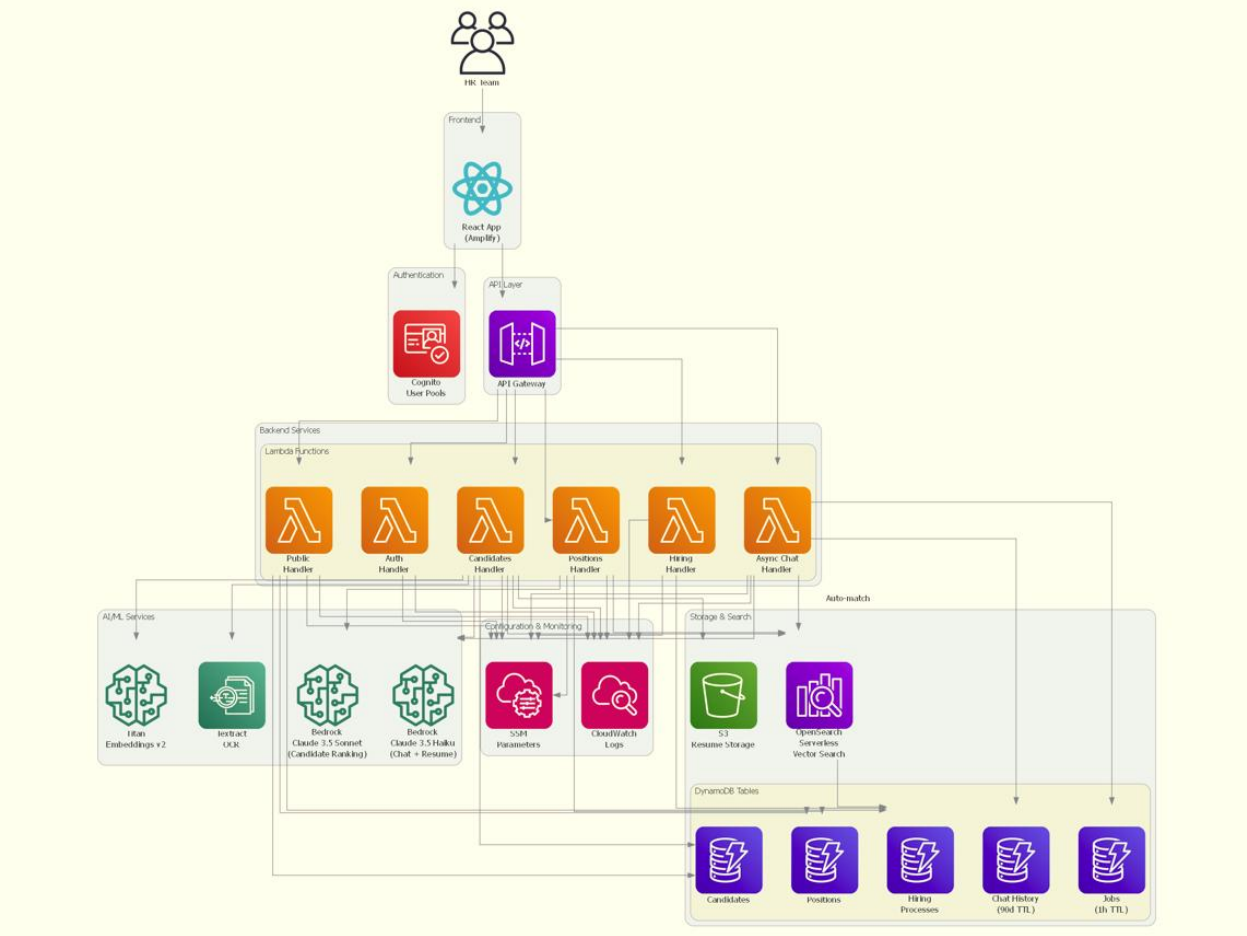
The orchestrator confirms details before executing irreversible actions (creating positions, starting hiring processes), preventing errors while maintaining conversational flow. This design balances automation speed with human judgment on critical decisions.

The tool agent encapsulation provides a clear evolution path

As recruitment workflows grow more complex (multi-stage interviews, offer negotiation, onboarding coordination), individual tool agents can be promoted to full autonomous agents with their own LLM reasoning. The orchestrator's interface remains unchanged, enabling incremental sophistication without architectural redesign.

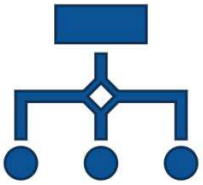
The serverless architecture on AWS ensures the platform scales with the growing talent pool without infrastructure management. Cost protection mechanisms (token limits, usage alerts, TTL-based cleanup) maintain predictable spending aligned with variable recruitment workloads.

IO Progression Talent Database High-Level Architecture



Next Steps

Potential areas for future enhancement of the agentic capabilities include extending the orchestrator to handle multi-turn offer negotiation workflows with autonomous counteroffers within defined parameters, enabling proactive candidate recommendations when positions are created without waiting for recruiter queries, adding a calendar tool agent for interview scheduling coordination, and extending the RAG agent to perform skill gap analysis across the talent pool against open positions.

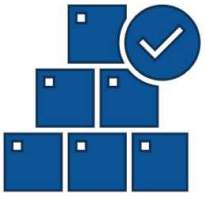


Intelligent Agentic Chat

This solution delivers context-aware, multi-turn conversational interactions using Claude 3.5 Haiku's native tool_use capability through Strands Agents SDK, enabling recruiters to execute recruitment workflows through natural language.

Secure and Responsible AI

Built with human-in-the-loop controls, prompt injection protection, session isolation, and AI output validation ensuring the orchestrator never fabricates candidate data and always confirms before irreversible actions.



Extensible Architecture

Designed to grow with complexity. Tool agents can be promoted to autonomous agents with their own LLM reasoning without modifying the orchestrator, enabling incremental sophistication as recruitment workflows evolve.

About Nova

Nova is a company specializing in Information Technology Consultancy Services. All our team members have one thing in common: our enthusiasm for technology and our passion for customer service excellence. We provide services in all North America, LATAM and Europe. Our headquarters are in NYC metropolitan area, and we also have offices in Guadalajara, Mexico and Madrid, Spain.

5