

Client Profile

Brito Ballesteros is a leading legal firm in comprehensive legal services and corporate consulting, with a consolidated track record in managing complex litigation and strategic advisory for the public and private sectors in Colombia. As a high-level professional services entity, its business model has been built on the precise interpretation of vast legal knowledge assets and adherence to rigorous regulatory standards.

To drive its technological evolution, the firm prioritized the implementation of Agentic AI architectures on AWS to automate its core legal operations and manage high-volume litigation cycles, transforming its unstructured information repositories into an intelligent, actionable knowledge base. This adoption allowed them to optimize legal research cycles by 85%, ensuring superior operational agility and data-driven decision-making.

- **Engagement Start Date:** December 15, 2025.
- **Engagement End Date:** Ongoing (Minimum 6-month continuous Managed Services contract executed and active).

The Business Challenge

Brito Ballesteros faced a critical limitation in its capacity for lead generation and operational response due to total reliance on the manual availability of its partners and legal team, a situation that became critically aggravated during peak litigation periods. By not having a 24/7 attention system capable of qualifying prospects in real-time, the firm experienced a constant loss of business opportunities (lead leakage) during high-traffic windows. Furthermore, managing medium and low-priority clients consumed valuable time from lawyers in administrative tasks of follow-up and document collection, delaying the conversion cycle and affecting billable hours.

Given this scenario, the need arose to implement an Agentic AI infrastructure on AWS that would function as an autonomous "triage brain" capable of managing massive demand peaks. The technical requirement consisted of a system capable of orchestrating consultative conversational flows on WhatsApp and Web, using Amazon Bedrock to perform expert pre-qualification based on complex business rules (analysis of debt amounts, income stability, and judicial status).

Finally, the technological need extended to Document Intelligence to eliminate the bottleneck in the discovery phase using Amazon Textract for the automatic extraction of data from sensitive legal documents. **Crucially, because this application became core to the firm's daily revenue generation, Brito Ballesteros required a Next-Generation Managed Service Provider to guarantee 24/7/365 operational stability, security compliance, continuous cost optimization, and proactive infrastructure monitoring.**

The Technical Solution

The technical architecture of the solution implemented and managed for Brito Ballesteros was based on an ecosystem of native AWS services, designed to offer a consultative, automated, enterprise-grade response:

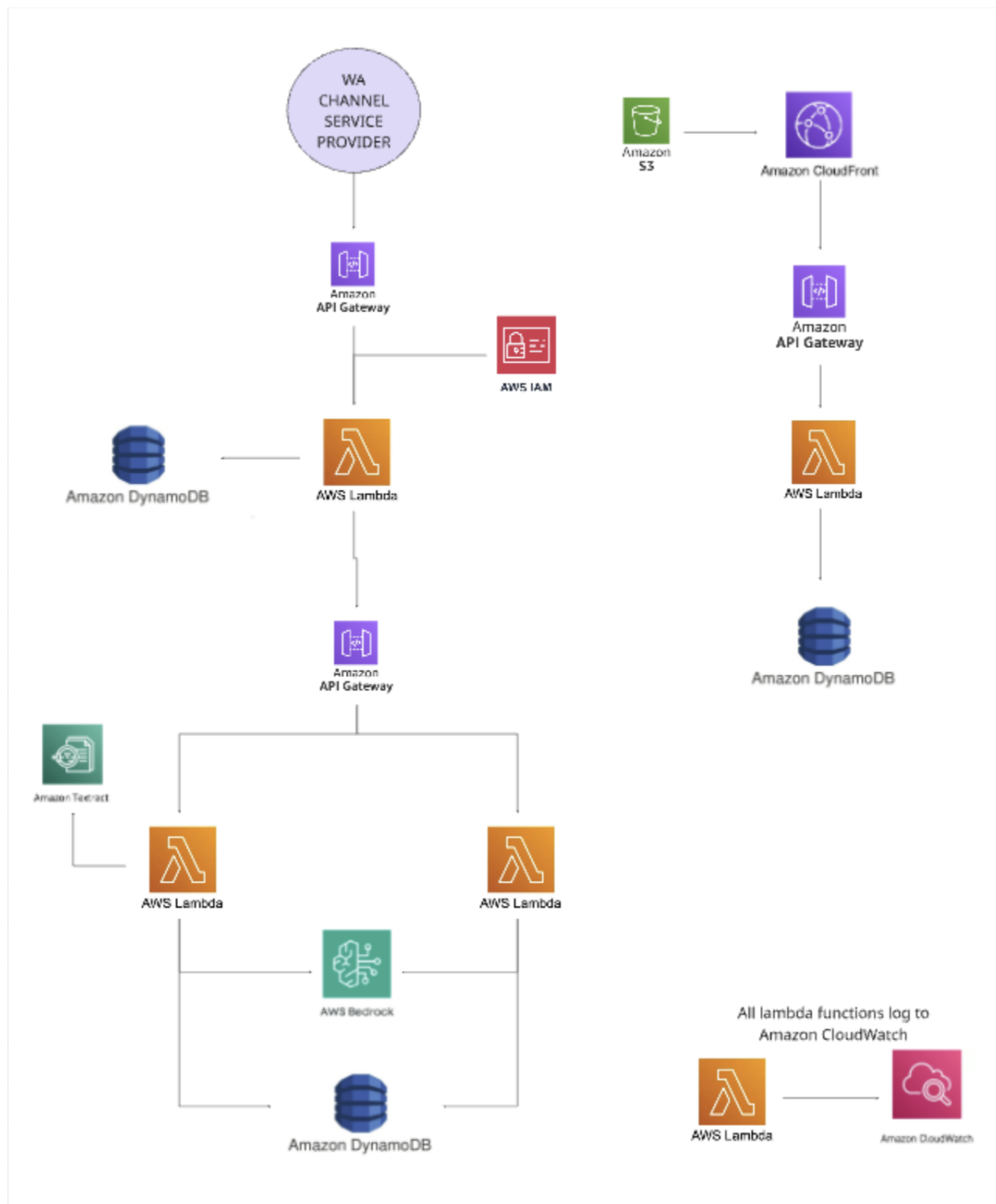
- **Core AI Brain:** Amazon Bedrock orchestrating Anthropic Claude 3 models, utilizing Chain-of-Thought reasoning to evaluate the legal merits of conversations and simulate preliminary consultations.
- **Document Intelligence:** Amazon Textract for automated data extraction from certificates of freedom, property titles, and default extracts.
- **Compute & Orchestration:** Serverless orchestration via AWS Lambda functions validating real-time documentation against prospect identities.
- **Data & Persistence:** Amazon DynamoDB for live conversational session state and Amazon S3 for secure document asset storage.
- **Security & Governance:** Document assets in Amazon S3 are protected by least-privilege IAM policies, AWS KMS encryption, and Amazon Bedrock Guardrails to prevent non-compliant legal advice. Amazon S3 lifecycle policies were implemented for the secure automated disposal of sensitive data after validation.
- **Observability & Management:** Centralized log and metrics collection via Amazon CloudWatch to monitor massive model invocations and system latency.

MSP Support and Managed Services (Day 2 Operations)

To ensure the continuous health, cost-efficiency, and evolution of the production environment, biia delivers comprehensive next-generation managed services covering the following scopes:

- **24/7/365 Operations & Service Desk:** biia provides continuous infrastructure monitoring and incident management via our multi-channel Service Desk. All automated notifications and critical system alerts are integrated directly into our ITSM platform. High-priority incidents—such as latency spikes in Amazon Bedrock or API failures—are managed under strict SLAs to guarantee zero disruption to the firm's legal triage.
- **Continuous Cloud Financial Management (FinOps):** Given that Amazon Bedrock and Amazon Textract utilization represents 75% of the total architectural spend, biia performs monthly cost optimization reviews. We actively manage Bedrock Provisioned Throughput during peak litigation cycles and enforce strict cost-allocation tagging to provide the client full visibility over their ROI and cloud budget.
- **Proactive Patching and Security Posture:** Under the AWS Shared Responsibility Model, biia manages and maintains the end-to-end security posture of the workload. This includes the automated auditing of IAM permissions with IAM Access Analyzer, ongoing vulnerability management, and maintaining automated backup routines in compliance with the firm's strict judicial data sovereignty policies.

Architecture Diagram



Quantitative Results

The implementation and ongoing management of the Agentic AI infrastructure generated a measurable transformation in the firm's operational efficiency:

- **Operational Agility:** 85% reduction in time dedicated to preliminary legal qualification, reducing initial triaging from an average of 5 business days to a 15-minute automated session in production.
- **Lead Capture:** 100% of high-priority leads receive an immediate response, eliminating lead leakage and increasing capture capacity by 40%.
- **Technical Accuracy:** Achieved 96% effectiveness in automatic data extraction via Amazon Textract and 92% accuracy in classifying 13 critical legal variables via Amazon Bedrock, ensuring an error margin of less than 1%.
- **Cost Optimization:** 80% optimization in infrastructure and technical support costs compared to traditional legal management models. Running on a 100% serverless architecture managed by biia reduced the operational cost per qualified lead by 50%.
- **Business Scale:** The production infrastructure operates with an estimated AWS Annual Recurring Revenue (ARR) of \$48,000 USD.

biia Profile and Added Value

biia positioned itself as a cutting-edge strategic partner specializing in the deployment and long-term management of Agentic AI ecosystems on AWS. Our value proposition focuses on a "Security by Design" architecture aligning with the highest standards of the AWS Well-Architected Framework. Through our Next-Gen Managed Services practice, we ensure that Brito Ballesteros can scale its consultative capacity securely, processing thousands of interactions with 99.9% architectural stability, zero operational friction, and continuous financial efficiency.

The success of this deployment has led Brito Ballesteros to designate biia as its exclusive AI Innovation and Managed Services Partner for future expansions into automated judicial monitoring.

Consent & Approval Note: This case study is published with the express written permission of Brito Ballesteros. The client has reviewed and approved all performance metrics, technical architecture details on AWS, and sustained operational results for inclusion.



About biia

biia is a premier next-generation technology consulting firm and cloud services provider specializing in the architecture, deployment, and long-term managed operations of Generative AI ecosystems, serverless applications, and high-scale data solutions on AWS. Committed to driving exponential efficiency and strategic agile innovation, biia transforms traditional workflows into autonomous, enterprise-grade digital platforms aligned with the highest standards of the AWS Well-Architected Framework. Through its Next-Gen Managed Services practice, the firm delivers robust 24/7 proactive monitoring, full-stack observability, cloud financial governance (FinOps), and rigorous security architectures, enabling organizations worldwide to operationalize technical excellence and scale business outcomes seamlessly in the cloud.