



Case Study: AI-powered Web scraping Agent using AWS Bedrock Agentcore

PROJECT OVERVIEW

An AI-powered Web Scraping Agent built on AWS Bedrock AgentCore that automatically reads, navigates and retrieves content from a target website to answer user questions with a full source citation.

PROBLEM STATEMENT

Organisations waste time manually searching websites and answering repetitive questions, while users struggle to find accurate information which leads to delays and increased support workload

- Manual ticket handling and repetitive query resolution by support teams
- Scattered site content, which requires users to search multiple pages
- Heavy reliance on support workers for questions that could be self served
- Risk of inaccurate answers when staff respond without checking the source
- Limited visibility into recurring support trends and common queries

SOLUTION OVERVIEW

Implemented an AI-powered Web Scraping Agent to automate website content retrieval and response generation

- Provides answers based on verified website content
- Provides context aware conversations for follow up queries
- Ensures isolated user sessions
- Automatically crawls and discovers relevant content across the site
- Cites the exact source page URL for every response provided
- Suggests three contextual follow up questions after every query
- Accessible via a simple chat interface embedded in AFIVE

METHODOLOGY

Step	Stage	Description
1	User Queries	User types any question in the agent chat window
2	Application Routes	AFIVE routes the question to the Web Scrape Agent
3	Agent Navigates	Agent reads the website, scores links and follows the most relevant pages up to 5 levels deep
4	LLM	Scraped content is sent to the LLM for response generation
5	Final Response	A clear, structured response is returned with the source URL cited and 3 follow up question

KEY FEATURES

Web Crawling

- Parallel scraping with up to 5 simultaneous page requests
- Query based link scoring, agent follows the most relevant pages first
- Configurable crawl depth, page limits, and content size caps
- Deduplication of URLs to avoid repetitive scraping

Two Model Architecture (Not fully implemented)

- For the questions which require scraping
 - Nova Micro will be used for content extraction
 - Claude Sonnet 4.5 will be used for reasoning
- For the follow up questions, which do not require scraping
 - Claude Sonnet 4.5 will be used for extracting the content (which is stored in Bedrock AgentCore Memory)
 - Nova Micro will be used for reasoning

Document Support

- Detects embedded document links on scraped pages
- Depending upon the client requirements, two approaches are implemented
 - The agent is programmed to read the docs and give the content and also it is programmed to just give the links to those docs and not read the content

Document Support

- Combined identity tracking via Device ID, Cookie ID, and IP
- For Guest Users
 - Once they reach 5000 tokens, CAPTCHA is returned and before email verification, the tokens are extended till 10,000
- For Logged Users
 - After Email Verification step, the tokens are extended till 50,000, once the user hits that limit, the user will have to wait for 24 hours to use the agent

TECHNICAL ARCHITECTURE

Component	Technology
AI Agent Runtime	AWS Bedrock AgentCore
Agent Framework	Strands Agents SDK (Python)
Large Language Model	Claude Sonnet 4.5 and Nova Micro
Conversation Memory	AWS Bedrock AgentCore Memory
Web Scraping	Python (BeautifulSoup)
Chat Interface	Gradio (Python)
Embedding Platform	AFIVE (Azure)
Secrets Management	AWS Secrets Manager
Network Security	AWS VPC
Monitoring	Amazon CloudWatch
Access Control	AWS IAM, JWT Tokens

SERVICE IMPLEMENTED

- AWS Bedrock AgentCore Runtime and Memory
- Claude Sonnet 4.5 and Nova Micro via Amazon Bedrock
- Strands Agents SDK with A2A protocol support
- Redis content caching including TTL
- AWS Secrets Manager for all configuration and credentials
- AWS VPC with private subnets and security groups
- Amazon CloudWatch for logging, tracing and monitoring
- Gradio chat interface

SECURITY & GOVERNANCE

- Access controlled IAM role, only authorised systems can invoke the agent
- JWT tokens used between AFIVE and the agent
- Agent reads only the public website content
- Each user session and conversation history fully isolated
- All credentials stored in AWS Secrets Manager
- Amazon Bedrock Guardrails to prevent inappropriate responses
- Rate limiting and token budget limits to prevent cost overrun

BENEFITS

Operational Benefits

- Reduced support team workload for common and repetitive queries
- Consistent and accurate responses sourced from live website content
- Full source URL citation on query
- Scalable to any public website with configuration changes

Business Benefits

- Enhanced user satisfaction through accurate self service
- Scalable support operations across departments and websites
- Enterprise grade security and governance
- Reusable solution deployable across multiple client websites

FUTURE STEPS

URL Caching

- If multiple users scrape same URL, it should return the cached result
- Vectorisation
- Planning to implement a separate DI agent, were the agent reads the doc content
- Planning to create a Multi Agent Orchestration
 - Webscrape Agent
 - DI Agent

Adactin is a leading global IT consultancy headquartered in Australia, specialising in Microsoft Solutions, Quality Engineering, Cloud Intelligence (AWS), and AI-powered Digital Engineering. With a strong presence across the APAC region, our expert consultants deliver enterprise-grade solutions that empower businesses to innovate and grow.

Trusted by government and enterprise clients alike, Adactin holds multiple multi-year panel contracts and has earned a reputation for excellence, reliability, and results. We are committed to helping organisations achieve operational efficiency, elevate customer experiences, and drive sustainable growth through smart, scalable technology solutions.

Contact us:

📞 1300 232 286

📍 Suite 402, 56 Station Street East Parramatta NSW 2150 Australia
