

Edward Budiman

+61 467643028 • ejbudiman@gmail.com • [linkedin.com/in/edward-budiman7](https://www.linkedin.com/in/edward-budiman7) • [edwardbudiman.com](https://www.edwardbudiman.com) • Sydney, Australia

SUMMARY

AI and Automation Engineer with hands-on experience architecting and delivering enterprise-scale **Generative AI, RAG**, and **Agentic AI** systems in production. Skilled in translating ambiguous business problems into secure, scalable, and reusable AI architectures across **AWS** and **Azure**. Experienced with **MCP** server integration, **multi-agent** protocol design, and full **Infrastructure-as-Code** deployment. Equally comfortable defining architecture standards, leading technical reviews, and communicating complex AI concepts to non-technical stakeholders and governance teams.

EXPERIENCE

NICE, Solutions Engineer, AI Interactions Engineer, CX June 2026 – Present

- Leading end-to-end AI solution implementations, from pilot to full deployment, designing scalable and high-performing conversational AI systems for enterprise customers using NICE CXone.
- Configuring and developing AI-powered applications including chatbots and voicebots, integrating NLU, ASR, IVR/ACD, and knowledge-driven capabilities with a focus on architectural rigor and system stability.
- Integrating enterprise systems across CRMs, RESTful APIs, and customer ecosystems to deliver cohesive end-to-end technical architectures that connect people, data, and workflows.
- Collaborating cross-functionally with customers and internal teams to translate business requirements into reliable AI solutions, supporting UAT, go-live readiness, and post-deployment optimization.

Kipanga, AI & Automation Engineer May 2025 – Present

- Architected and delivered a multi-surface enterprise RAG platform (Flask, FastAPI, Pinecone, AWS Lambda, Docker), surfacing company knowledge through a web chatbot, Slack bot, and an MCP server integrated directly with Claude Desktop via a stdio/SSE local proxy, one of the earliest production MCP server deployments in the Australian market.
- Designed and implemented multi-agent and agentic workflows with tool invocation, multi-step reasoning, contextual memory handling, and structured fallback logic, selecting frontier LLMs (OpenAI, Anthropic, Google) based on latency, context window, reasoning depth, and cost tradeoffs.
- Built reusable Infrastructure-as-Code via Pulumi across multiple AWS stacks (Lambda, API Gateway, ECS, EventBridge, DynamoDB, S3), including a serverless authenticated webhook scheduling system and a standalone Pinecone-backed context API.
- Implemented IAM policies and RBAC-based access controls across AWS, Azure, APIs, and data layers, enforcing secure and governance-aligned handling of sensitive information within AI systems.
- Identified and scoped high-value automation opportunities across internal and client workflows, translating ambiguous commercial problems into reliable, iterable AI solution blueprints.

Triputra Consulting, Data Scientist Jan 2024 – April 2025

- Developed AI-powered chatbots and knowledge-base systems using RAG pipelines, fine-tuned LLM workflows, and vector search to support internal teams and client-facing use cases.
- Built and automated data pipelines using SQL, Python, AWS Glue, Lambda, and S3, reducing processing time by 50% and improving reporting accuracy across business units.
- Partnered with cross-functional stakeholders to convert ambiguous commercial challenges into structured analytics and AI solutions, contributing to measurable revenue growth and operational efficiency.

Mido, Data Analyst Oct 2023 – Jan 2024

- Built a Tableau sales/inventory dashboard automating reporting for 5+ vendors, improving data accuracy by 30% and reducing manual workload.
- Conducted 5-year fashion industry trend analysis informing vendor promotional strategies that boosted seasonal sales by 25%.

Dewaweb, Data Analyst Jul 2023 – Sep 2023

- Analysed customer behaviour data (Python/R) to refine e-commerce marketing strategies, improving affiliate conversion rates by 18%.
- Supported vendor onboarding with cloud-based analytics tools, reducing data processing delays by 40%.

SKILLS AND CERTIFICATIONS

- **Certifications:** AWS Cloud Practitioner (2025), AWS AI Practitioner (2025), AWS Developer Associate (2026), NiCE Cognigy - Associate - NCAE (NiCE Certified AI Engineer) (2026), Make.com
- **AI & Generative AI:** RAG pipeline design, multi-agent and agentic workflow design, LLM fine-tuning, data flywheel concepts, prompt engineering and orchestration, MCP server integration, LangChain, LangGraph, HuggingFace Transformers, PyTorch, OpenAI, Anthropic, Google GenAI
- **Backend & APIs:** Python, Node.js (TypeScript), REST API design, WebSockets, SSE, Flask, FastAPI, Fastify
- **Cloud & Infrastructure:** AWS (Lambda, API Gateway, S3, Glue, SageMaker, IAM, DynamoDB, ECS, EventBridge), Azure (Container Apps, Container Registry, AD), Docker, Pulumi, GitHub Actions, CI/CD, Git
- **Data & Analytics:** SQL, Python, R, AWS Glue, Athena, Airflow, dbt, Snowflake, Power BI, Tableau
- **Frontend:** React 18, TypeScript, Tailwind, shadcn/Radix UI, TanStack Query, Vite
- **Automation Platforms:** Make.com, n8n, Zapier, Pipedream, Cognigy, CXOne

SELECTED PROJECTS

Enterprise RAG Knowledge System | Enterprise Client | Python, Flask, FastMCP, Pinecone, AWS Lambda, Docker

- Multi-surface RAG platform with a web chatbot, Slack bot, and MCP server on AWS Lambda. Ingests LinkedIn, Google Drive, and Wrike data into Pinecone vector stores. MCP server integrates directly with Claude Desktop via a stdio/SSE local proxy.

Meeting Minutes Automation | Enterprise Client | React 18, FastAPI, Node/Express, OpenAI, Microsoft Graph, Docker

- End-to-end system for generating formatted Word meeting minutes from raw transcripts. Includes a React 18 frontend, FastAPI backend, and an email-triggered Node/Express service using Microsoft Graph. Dockerised for deployment.

AI Chat App | Independent Client | Python, FastAPI, React 18, TypeScript, SSE, OpenAI, Anthropic, Docker, Fly.io

- Full-stack streaming chat application with provider-agnostic LLM switching (OpenAI/Anthropic) and a server-side tool-calling loop. Implements end-to-end token streaming via SSE with mid-stream UI affordances, a multi-turn tool loop bounded to 5 rounds, and a single tool registry that derives both provider-specific formats. Deployed on Fly.io with Vercel frontend.

Multi-Agent Research Assistant | Independent Client | Python, LangGraph, LangChain, OpenAI, Anthropic, Tavily

- Cyclic multi-agent graph built on LangGraph's StateGraph where a Supervisor routes between a Web Searcher, Summariser, and Synthesiser using structured LLM output, with deterministic guardrails against runaway loops. Streaming Synthesiser outputs per-token to the terminal. Pluggable architecture supports new workers, tools, or a MemorySaver checkpointing without touching the orchestration logic.

LLM Eval Harness | Independent Client | Python, Typer, Pydantic v2, OpenAI, Anthropic, deepeval

- Production-quality CLI for evaluating LLM outputs using the LLM-as-judge pattern. Provider-agnostic client abstraction, strict-JSON judge contract with explicit safeguards for verbosity bias and calibration drift, and optional deepeval integration for relevancy, faithfulness, and hallucination metrics. Pydantic v2 models enforced across all public surfaces.

EDUCATION

BSc in Computing, Majoring in Computational Data Science | University of Sydney

- Awarded 20% merit scholarship; coursework in business analytics, systems design, and cloud computing.