

SPARLI AMAN

Agentic AI Systems Engineer · Forward Deployed / Solutions Engineer

Full EU overlap, 4h US East | Islamabad, Pakistan (UTC+5) | Open to relocation (US / EU / UK) and customer travel
sparli@croncore.com | linkedin.com/in/sparli-aman

I design and build agentic systems that act on production infrastructure: tool-using agents wired into CRMs, ticketing, repos and regulated comms systems, with approval gates so nothing irreversible happens without a human. Seven years of client-side delivery, end to end. I scope with the customer's engineers, architect it, build it, deploy into their environment, and own it after go-live.

EXPERIENCE

Croncore, AI engineering consultancy · Co-Founder & CTO · Jan 2025 – present

Shurq, UK adtech startup, joined at idea stage · Founding AI & Data Engineer · Jul 2022 – present, advisory since 2025

Freelance AI / ML Engineer · Top Rated on Upwork, Level 2 on Fiverr · 2019 – 2022

SELECTED PROJECTS

Tool-using agent operating on live client systems

2025

Croncore internal delivery platform, in daily production use

Agentic tool-use system on the Anthropic Messages API. One tool surface, two runtimes: the same tools run against the API directly or through the Claude Agent SDK, so the agent can be driven from a server or from a developer's machine without rewriting the tools. It triages GitHub pull requests, assigns and tracks developer tasks, and drafts client status reports and email.

Every outbound action sits behind a draft-and-approve gate. The agent has no code path that posts to ClickUp or sends client mail directly. Writes land in a `pending_actions` table and wait for explicit release. Every run is scoped to one `project_id`, so context and memory cannot leak between clients. Model selection is per task rather than one model for everything.

Anthropic API · Claude Agent SDK · Next.js · Drizzle/Postgres · ClickUp, GitHub, Gmail APIs

Bilingual education platform, live with Québec school boards

2024 – present

Co-led delivery; architecture owner for the next stage

Co-led the EN/FR role-based platform now in use with Québec school boards. Django/DRF, Next.js/React 19, PostgreSQL, a Gemini image-generation pipeline for lesson assets, JWT auto-refresh behind an HTTP/HTTPS proxy layer, and an EN/FR-parity Playwright suite so a regression in one language cannot ship because the other passed.

Lesson images were rendering broken text. I rewrote the generation prompts and the text handling around them, which cut **text-rendering errors by roughly 80% with no model change, no added cost and no added latency**, checked against before-and-after output on real lesson content.

Now extending it. I designed an orchestration layer where the content, assessment and analytics agents read and write one shared learner record in Postgres, so a tutoring decision can use what the learner did in earlier sessions instead of starting cold. I wrote the specification the team builds from, **picked the team, and build alongside them.**

Django/DRF · Next.js/React 19 · PostgreSQL · Gemini · Playwright · JWT/proxy auth

Multi-provider LLM platform under load

2023 – 2025

Shurq, UK adtech startup

Production LLM API platform with tiered pricing and automatic provider fallback, self-hosted Ollama through to OpenAI, so a provider outage degraded cost and latency instead of taking the feature down. **Stress-tested to 200+ concurrent users** before release, with latency and drift dashboards handed to the team so degradation shows up on a graph before a customer reports it. Built and led the data-engineering team around it.

Python · Ollama · OpenAI · PostgreSQL · BigQuery · Prometheus/Grafana · GitHub Actions · AWS

Two clients: US payment services · Canadian SEO SaaS

Conversational agents with real tool surfaces: pre-call enrichment from Google Places, full Pipedrive CRM integration (dedupe, deal creation, callback scheduling) as agent-invoked actions, and a custom SMS tool that delivers a document **mid-call** without breaking conversation flow. Built the compliance layers in-house: timezone-aware dialling windows, DNC scrubbing, mid-call opt-out detection, verbal-consent logging and CASL unsubscribe handling.

All the tuning came from live calls. **Swapped the conversational model from gemini-2.0-flash to claude-haiku-4.5 to fix response latency.** One failure took a while to find: speech-to-text was mishearing "hello" as "no", which tripped the end-call rule and hung up at four seconds. Fixed with never-end-on-first-turn and graceful-close rules, then wrote the spam-mitigation playbook the client still operates from.

Node.js · Twilio · ElevenLabs Conversational AI · Claude Haiku · GPT-4o-mini · Pipedrive · Render

Retrieval agent that quotes prices from the client's own data

2025 · automotive client

Took a Django prototype to a production retrieval pipeline: LangChain and LlamaIndex over pgvector with an LLM query-rewriter, and pricing synced live from the client's own sheets. A price change on their side lands without a redeploy, and no quoted figure comes from model memory.

Speech service into a legacy PBX

2024

Built and operated the serving stack for a language no commercial vendor offered. A self-hosted FastAPI service running two off-the-shelf speech engines behind one interface with fallback between them, producing audio that matched the client's existing Asterisk/FreePBX contract, so nothing on the telephony side had to change.

SKILLS

Languages	Python · TypeScript/JavaScript · Node.js · SQL
Agentic & LLM	tool-using agents · MCP tool servers · Anthropic tool use · OpenAI function calling · schema-validated structured outputs · RAG · provider fallback · human-in-the-loop approval gates · prompt-injection containment · eval design and calibration
Backend	FastAPI · Django/DRF · Next.js · PostgreSQL · pgvector · Redis · BigQuery · Docker · async pipelines · rate limiting and backoff
Cloud & ops	AWS · GCP · Azure · Vercel · Render · Cloudflare · GitHub Actions CI/CD · Prometheus/Grafana
Integrated with	Twilio · Asterisk/FreePBX · Deepgram · ElevenLabs · Pipedrive · ClickUp · GitHub · Gmail · Google Sheets · Google Places · Stripe <i>(each wired up and debugged in production)</i>

EDUCATION

BSc Software Engineering, Pak-Austria Fachhochschule, Institute of Applied Sciences and Technology, Pakistan

AVAILABILITY

Remote, with working hours overlapping the EU in full and the US East coast by four hours. Set up to work remotely long term, and equally open to relocation to the US, EU or UK where a role calls for it. Available for customer travel and on-site work.