Jae Lee

AI Solution Architect

Singapore, Singapore | LinkedIn

AI solution architect and technology founder with 20+ years building software products, platforms, and engineering teams across Asia, the Middle East, and the US. Specializes in applied AI, generative AI, AI agents, and enterprise system architecture.

Experience

Entrepreneur in Residence Oct 2025 — Dec 2025
Antler *Riyadh, Saudi Arabia*
Selected for Antler's Riyadh residency focused on founder formation and venture development. Co-founded FMInsight, an AI agent platform for facility management in Grade A commercial buildings across the GCC. Defined product strategy, market positioning, data models, AI use cases, and go-to-market plan for enterprise building operators.

Investor Jan 2025 — Jan 2025
Namecoach
Backed by Founders Fund, Metaplanet Holdings, Impact America Fund, Authentic Ventures

Investor Jan 2025 — Jan 2025
BackerKit
Backed by Sahil Lavingia, SV Angel, Y Combinator, Garry Tan

Digital Director Nov 2024 — Sep 2025
MBLM DXB *Dubai, United Arab Emirates*
Led applied AI commercialization and digital product delivery for enterprise and public sector clients across the GCC and Korea. Commercialized allumni.ai, MBLM's internal Generative AI platform, into client-specific workflows and demonstrations. Designed over 25 AI agent templates and led digital delivery spanning websites, CRM, analytics, and multilingual content operations.

Interim CEO Sep 2024 — Dec 2024
Madison Technologies *Singapore*
Provided interim executive leadership during organizational change. Clarified business priorities, aligned stakeholders, and brought direction across leadership, technology, Salesforce workflows, and company execution.

Startup Mentor Jan 2024 — Present
Wavesparks *Singapore*
Mentor founders on product strategy, technical architecture, go to market, fundraising readiness and early execution. Help teams narrow the problem, test assumptions, shape an MVP and turn broad ideas into clear next steps.
Also support Wavesparks venture building programs and workshops, including founder mentoring and structured innovation work for programs such as UBS LEADS and the Young Founders Summit. Wavesparks materials describe your mentoring focus as product strategy, technical architecture, go to market execution and fundraising.

Guest Lecturer Feb 2023 — Feb 2023
James Madison University *Harrisonburg, Virginia, United States*
Guest lecturer and workshop speaker on entrepreneurship, career development and startup building, drawing on practical lessons from Kempus and more than two decades of building technology companies.

Co-Founder and CEO Aug 2022 — Sep 2024
Kempus *Remote*
Co-founded and led a verified student community platform focused on anonymous participation, peer knowledge, and career readiness. Raised $3 million in seed funding and expanded to 33 universities across the United States. Led product, engineering, partnerships, and company strategy. Later integrated RAG and generative AI for personalized career guidance based on students' coursework and skills.

Panelist Apr 2022 — Apr 2022
W.Media - Global
Panelist at Digital Week South Asia 2022, discussing multi cloud and hybrid cloud strategy, cost management and security for enterprise technology environments.

Panelist Nov 2021 — Nov 2021
Horasis
Panelist at the Horasis Asia Meeting, discussing how AI and emerging technology could reshape Asian economies, organizations and markets following the pandemic.
Agenda: Economies are pressed by the COVID mutations. How will the new Asian revolutionary technology - AI and so on - impact the Asian world which hosts massive populations? How will their economies, geopolitics and organizations change? What technologies might rise to the fore?

- Bernard Moon, Co-founder and Partner, SparkLabs Group, USA

- Clas Neumann, Senior Vice President, SAP, China
- Jae Lee, Vice President of Engineering, Quincus, Singapore
- Sudhir Sharma, Chairman, Mitkat Advisory Services, India
- Chaired by Benjamin J. Butler, Co-Founder, Quantum Futures, Hong Kong

Panelist Nov 2021 — Nov 2021
PCCW GLOBAL Limited
Panelist for a PCCW Global discussion on interconnectivity in APAC, covering cost effective cloud network design, connectivity strategy and infrastructure best practices.
Agenda: Creating Cost Effective Cloud Network Solutions and Best Practices

Vice President of Engineering Jan 2021 — Aug 2022
Quincus *Singapore*
Led a distributed engineering organization of 70+ developers across 8 countries for supply chain visibility and logistics orchestration products. Directed architecture and delivery across API gateways, middleware, data transformation, and customer integrations handling ~800 million API calls per month. Led applied AI capabilities in dynamic rating, forecasting, and routing. Oversaw integrations with ERP, WMS, OMS, CRM systems and carriers including UPS.

Investor Apr 2020 — Dec 2024
FC Pinzgau Saalfelden

Guest Lecturer Oct 2019 — Oct 2019
National University of Singapore *Singapore*
Guest lecturer for the NUS Business School entrepreneurship module, sharing practical lessons from building startups, launching digital products and leading technology teams.

Startup Mentor Sep 2019 — Present
Founder Institute
Mentor early stage founders across product strategy, technical architecture, pitch clarity, fundraising readiness and go to market. Help founders test assumptions, make practical product choices and move from idea to early execution without adding unnecessary complexity.

Official Member Sep 2019 — Sep 2026
Forbes Technology Council
Contributed expert commentary through Forbes Technology Council on platform engineering, application performance, data analysis, AI readiness, software team practices, retail technology and connected customer experiences.
Shared practical lessons drawn from building software products, leading engineering teams, scaling platforms and applying AI across real business settings.

Startup Mentor Sep 2019 — Sep 2020
ACE.SG (Action Community for Entrepreneurship) *Singapore*
Mentored Singapore startups on product direction, technology choices, go to market, fundraising readiness and founder decision making. Focused on practical steps that teams could test and execute quickly.

Guest Lecturer Sep 2019 — Sep 2019
Nanyang Technological University Singapore *Singapore*
Guest lecturer on entrepreneurship and technology, drawing on founder, product and engineering experience across startups and international digital businesses.

Chief Technology Officer Dec 2018 — Dec 2020
WorldRoamer® *Singapore*
Led technology strategy and built the technical foundation for an online travel platform serving travelers across Asia and the United States. Owned platform architecture, product engineering, travel data integration, and security. Integrated CRS platforms and channel managers to bring 50,000+ travel listings to MVP. Implemented machine learning-based fraud detection and transaction risk workflows.

Vice President of Engineering Jul 2017 — Sep 2018
RingMD *Singapore*
Led engineering for a telemedicine platform spanning live video consultation, patient and provider workflows, Salesforce CRM, healthcare chatbots, and wearable IoT integrations. Supported RingMD through Singapore Ministry of Health's telemedicine regulatory sandbox and expansion through India's Common Services Centre network.

Director, Head of Creative Technology Lab | Creative Technology Director Jun 2016 — Jun 2017
Group IDD *Seoul, South Korea*
Led creative technology, digital product, and campaign technology work for major brands including Samsung, Hyundai, Kia, L'Oréal, Burberry, and others. Directed work across web, mobile, AR, VR, messenger chatbots, social analytics, live streaming, and location services. Notable projects included Samsung Galaxy Tab S3 digital launch, L'Oréal influencer analytics, and Burberry Web VR. MEET for airberlin placed third in Best Business Experience category at global Flight to Innovation hackathon.

Advisor Jan 2016 — Present
Tech Ranch *Austin, Texas, United States*
Advise founders on product, technology, market entry and startup growth. Supported international founder programs and cross border market entry work, including helping Korean startups understand and enter the U.S. startup market.

Founder and CEO Jan 2013 — Dec 2016
MovableAds *Remote*
Founded a sports technology and advertising startup focused on mobile fan engagement and data-driven marketing. Led strategy, product development, software delivery, fundraising, partnerships, and market testing. Developed mobile experiences that captured live fan reactions to help teams and advertisers understand audience interests.

Lead Software Engineer Jan 2012 — Dec 2013
Cheil Worldwide *Seoul, South Korea*
Led software engineering and technical assessment for Samsung and Cheil digital initiatives across mobile products, web applications, APIs, campaign platforms, and developer programs. Worked with creative and strategy teams to deliver technically sound digital experiences. Supported Samsung developer ecosystem work and global mobile application initiatives.

Tech Lead Jan 2011 — Dec 2012
XenoOne *Seoul, South Korea*
Led technology development for LifeCruz, a social lifelogging platform that organized personal social content using context like location, time, and hashtags. Led architecture and API integration across Facebook and Twitter. Built product and engineering foundations for a startup later acquired by Cheil Worldwide.

Co-Founder and CTO Jan 2010 — Dec 2011
JellyCoaster *Seoul, South Korea*
Co-founded JellyCoaster and led mobile product development for consumer applications and branded digital products. Built location-aware, mapping, and social experiences during early smartphone and App Store era. Led development of Finger Map, which received U.S. App Store editorial recognition, and OPTIMAP, a location-based mobile experience for Kia Motors.

Founding Lead Developer Jan 2009 — Dec 2010
BRCH *Seoul, South Korea*
Led architecture and development for e-commerce, logistics, order management, and freight forwarding systems. Built Modubuy.com as a cross-border commerce platform connecting eBay auctions, payments, warehouse workflows, and delivery. Developed a live auction engine with automated bidding and concurrent auctions. Created a bilingual freight forwarding portal with shipment tracking and quotation workflows.

University Lecturer Jan 2007 — Dec 2009
University of Jendouba *Béja, Tunisia*
Taught university-level coursework.

Lead Developer, Contract Jan 2006 — Dec 2007
Oneday Mall *Seoul, South Korea*
Led development for e-commerce platform.

Founder and Lead Developer Jan 2004 — Dec 2005
Jayster Network *Seoul, South Korea*
Founded and led development of network application.

Digital Producer, Contract Jan 2003 — Dec 2004
Blooming Daisy Production *Seoul, South Korea*
Produced digital content and projects.

Web Developer Jan 1999 — Dec 2002
Suntech Electronics *Incheon, South Korea*
Developed web applications and services.

Co-Founder Jan 1999 — Dec 2000
K.E.L. Korea Engineering Lab *Seoul, South Korea*
Co-founded engineering lab.

Education

University of Illinois Urbana-Champaign
Business Administration
Non-Degree Program

HEC Paris
Executive Education

Pursued studies in Executive Program, discontinued program for professional development.

성균관대학교(SKKU)
Master of Business Administration - MBA
Academic scholarship recipient; coursework completed, thesis pending

Cornell University ILR School
Management Certificate

Kyung Hee University
Bachelor of Engineering
Received Academic Scholarship. Pursued studies in Computer Science Engineering before discontinuing the program to focus on professional development and launching a startup venture.

Hanyang Cyber University
Bachelor of Engineering, Computer Science Engineering

Projects

Managed multi-model AI inference platform
Built and hardened a managed, multi model AI inference platform with an OpenAI compatible API, provider routing, verified model catalogs, and centralized operational controls.
The platform brings together Clerk authenticated workspaces, tiered Stripe subscriptions, rolling SGD credit accounting, usage tracking, API key management, and administrator controls. It also supports Telegram bots with one configured model per bot, isolated conversation history, reset controls, image input, voice transcription, streaming responses, managed webhooks, and bounded reply delivery.
Web search is available through user owned Exa or Brave credentials, with encrypted storage, provider verification, exact URL retrieval, hostname restricted fallback search, failure safe handling, and only one active provider per user. The application also includes English, Simplified Chinese, and Malay localization.
Security and reliability controls cover shared inference quotas, concurrent provider work, expiring database leases, browser access boundaries, pinned HTTPS transport, normalized image handling, account suspension, subscription grant idempotency, and safe migration replay.
The result is a practical platform for serving multiple AI models through one consistent API while maintaining control over access, billing, usage, security, and administration. Internally, the service runs through a controlled local compute environment based in Singapore for operational testing and validation.

Full stack AI agent workspace with reasoning and browser automation, Agent Orchestrator
Built a full-stack AI agent workspace with streamed conversations, visible reasoning, and provider/model management. Implemented per-conversation browser sessions, URL checks, encrypted credentials, and Playwright-based browser tools for web research and automation.

Multilingual voice interpretation prototype for physiotherapy, Chat4Physio
Built a working prototype for physiotherapy consultations with speech transcription, translation, audio playback, and terminology controls. Includes consultation records, exercise video access, and multilingual summaries.

Applied AI prototype for community reporting on child safety, Guardian Watch
Built a working applied AI prototype for community reporting with multimodal intake, AI-assisted classification, severity assessment, location extraction, and pattern detection. Includes controlled access, auditability, and human-led review workflows.

JustRelay - AI home services concierge for Expats in Korea
Living as a Western expat in Korea means navigating a language barrier for even the most routine home service needs; finding a plumber, booking a cleaner, or scheduling a handyman requires Korean fluency most expats don't have. JustRelay solves this with a single English message: describe your need, and the AI concierge handles everything else.
The product is built around a conversational AI agent that matches users to a curated, vetted directory of Korean service providers, ranks them by relevance, proximity, and review strength, then either submits appointment inquiries directly on the user's behalf (via web forms) or generates ready-to-send Korean messages, politely worded in natural honorifics, that the user can copy and paste into KakaoTalk, SMS, or email. The agent handles scheduling logic, avoids double-booking, and tracks each request through to completion.
On the technical side, JustRelay is a full-stack TypeScript application in a pnpm monorepo: a React + Vite + Tailwind v4 frontend, an Express 5 API server backed by PostgreSQL via Drizzle ORM, and an agentic LLM layer built on OpenRouter with a contract-first OpenAPI spec driving all client/server type safety through Orval codegen. The admin console includes a business directory with AI-generated review snapshots and sentiment analysis, a GA4 analytics integration, a markdown content CMS, and a research assistant agent for growing the provider directory. Authentication is session-based with email verification via Resend.

Agent CMO
Agent CMO is an AI-native marketing operating system designed to transform campaign briefs into structured strategy, reviewed creative assets, performance hypotheses, and iterative learning loops. The platform is built around a master-agent

orchestration model that coordinates specialized sub-agents across brief intake, market research, ICP definition, positioning, channel planning, content strategy, copywriting, creative direction, paid media, email sequencing, QA, CRO, performance forecasting, memory, and experimentation.
The core architecture uses a TypeScript monorepo with a React, Vite, Tailwind, and Wouter frontend, an Express and Node.js API layer, and a PostgreSQL database managed through Drizzle ORM. Agent pipeline jobs are persisted in the database and executed through an in-process worker, enabling campaign stages to run in a controlled sequence with transparent status tracking, structured outputs, and recoverable execution. Shared packages support schema consistency, generated API contracts, provider integrations, authentication, storage routing, and reusable campaign data models.
To make the system useful beyond one-off generation, Agent CMO implements durable campaign memory and experiment feedback loops. The Memory Agent stores brand voice, audience insights, approved learnings, and prior campaign context, while the Experiment Agent converts performance signals into new hypotheses for future campaigns. This allows the product to behave less like a generic AI content tool and more like a coordinated marketing team that can plan, produce, review, remember, and improve.

Secure AI file intelligence and RAG platform, SourceMind
Built a secure AI file intelligence platform with OAuth 2.0 PKCE, AES-256 GCM encryption, and hybrid search. Supports source-grounded answers with citations using PostgreSQL, pgvector, and ranking fusion.

Persistent memory server for AI agents using MCP and PostgreSQL, Agent Memory Dump
Built a persistent memory server connecting AI agents to stored information using Model Context Protocol and PostgreSQL. Designed seven MCP tools with hybrid retrieval combining full-text search, vector similarity, and recency weighting, plus local ONNX embeddings and token-aware retrieval.

AI SaaS product generating personalized manga-style webtoons, Memtoon - AI Webtoon Generator
Built and deployed an AI SaaS product that converts recent life events into personalized manga-style webtoons using a multi-model generation pipeline. Selected as a Judge's Favorite at the Replit Agent 4 Buildathon.

MCP server connecting AI agents to Odoo ERP, Odoo MCP
Built an MCP server enabling AI agents to interact with Odoo ERP. Supports structured data access, CRUD operations, workflow execution, model introspection, and development/demonstration data workflows.

Cofounder GPT
Developed a custom GPT designed to help early-stage founders stress-test startup ideas and pitches for problem-solution fit.
Cofounder GPT acts as a structured thinking partner — not a cheerleader. It runs every idea through a defined diagnostic flow: restating the concept, scoring the problem using the FUGUPP framework (Frequent, Urgent, Growing, Unavoidable, Painful, Popular), rewriting vague problem statements in solution-free language, identifying the top reasons a pitch will fail, and producing a lean validation plan with testable hypotheses, customer interview questions, and prototype options with pass/fail criteria.
Built on advanced prompt engineering with embedded design thinking, lean startup methodology, and SCAMPER-based ideation flows. Optimized for founders who need clarity and speed over encouragement.

Dubai Metro API
The Dubai Metro Stations API provides structured access to station-level information for Dubai's rail network, including Dubai Metro and Dubai Tram stations. It includes station names, line associations, geolocation data, and ingestion metadata.
This API enables developers to build applications for transit discovery, location-based services, urban mobility tools, travel planning, and rail network analysis by offering flexible filtering capabilities and clean, structured responses.
Key Capabilities:
• Search stations by station_name and line_name • Perform partial and wildcard matching for flexible queries • Retrieve detailed metadata including: • Station name • Line name • Nested geolocation object with longitude and latitude • Loading and ingestion metadata

Professional Baseball Analytics RAG Assistant
Developed a Retrieval-Augmented Generation (RAG) conversational agent designed to analyze and query comprehensive historical statistics for a professional baseball league. The system empowers users to interact naturally with complex datasets, transforming raw player metrics, situational statistics, and team histories into easily digestible insights. By leveraging modern AI design patterns, the project successfully bridges the gap between rigid database queries and natural language understanding.
At its core, the application features a robust backend powered by a locally hosted large language model, ensuring complete data privacy and low latency without relying on external APIs. The technical architecture includes a custom search index and an optimized RAG pipeline that accurately retrieves relevant context before formulating a response. Additional technical highlights include automated text romanization and translation modules to support multilingual queries, along with a dedicated analytics pipeline for advanced metric evaluations.
To maximize accessibility, the project offers both a lightweight terminal user interface and a fully featured graphical application. Users can seamlessly request career data exports, compare situational statistics, and pull specific player biographies, with the backend intelligently routing natural language requests to specialized tool handlers. This application

demonstrates a comprehensive capability to architect end-to-end AI solutions, integrating complex data pipelines, local model orchestration, and user-friendly frontend experiences.

Ripplet.me
Built an open-source, full-stack community discussion platform modelled after Reddit, with AI-powered moderation built in from the ground up.
Users can create forums, post discussions (Ripples), thread nested comments up to 5 levels deep, and react with emoji. A role hierarchy (User → Moderator → Admin → Super Admin) governs moderation, with shadow banning, post hiding, and a full audit log.
The stack is a pnpm monorepo: React + Vite + TypeScript + Tailwind CSS + shadcn/ui on the frontend; Express + TypeScript + PostgreSQL (Drizzle ORM) on the backend; OpenAPI 3.0 with Orval-generated React Query hooks and Zod validators shared across the workspace. Auth is Replit OIDC + PKCE with 7-day database-backed sessions.
The AI moderation layer connects to Anthropic Claude (Haiku, Sonnet, or Opus; selectable per forum). Moderators configure custom system instructions, a blocklist, an automated posting schedule, and can let the AI publish its own posts under the RippletAI identity.
Security hardening includes a tightened Content Security Policy, rate limiting, ReDoS-safe input caps, a URI allowlist in the HTML sanitiser, and parameterised queries throughout.
Fully open-sourced under MIT license at github.com/leejaew/ripplet.me and publicly accessible at the deployed Replit endpoint.

SG MRT Exit Finder
Built and launched SG MRT Exit Finder, a custom GPT on ChatGPT that enables search, navigation, and analysis of Singapore MRT station exits using LTA GeoJSON data. The project includes a self-developed API that serves MRT exit data via a secured endpoint with HTTP Basic Authentication, supporting flexible station name queries with wildcard capability.
The GPT integrates with this API through a structured OpenAI action schema, enabling real-time data retrieval directly within ChatGPT. Query interpretation and response formatting are handled entirely by the GPT based on user intent — a single API integration consolidates what would otherwise require multiple discrete tool definitions.
Core capabilities include exit search by station name, exit-level detail retrieval, Google Maps link generation, nearest-exit lookup by coordinates or landmark, radius-based exit discovery, intra-station exit comparison, and location-based queries for logistics, tourism, and urban planning scenarios.
Note on limitations: As a custom GPT, this solution is platform-locked to ChatGPT, restricted to OpenAI's action schema, and limited in how reliably it handles multi-step or complex spatial queries — the model interprets all tool behaviour dynamically rather than through explicitly defined, typed interfaces. For production workflows requiring consistent, auditable, and multi-step agentic reasoning across platforms, implementing a dedicated MCP server paired with an AI agent is the more appropriate architecture.

SG MRT Exits MCP
Built a production-ready Model Context Protocol (MCP) server that bridges AI agents with Singapore's LTA MRT Station Exit open dataset, enabling natural language and agent-driven queries across 597 exits and 186 stations.
The server exposes a standardized set of tools that any MCP-compatible AI client (such as Claude, Manus AI, Cursor, or custom agents) can call to answer real-world questions about Singapore's transit network — without manual API integration required on the agent side.
Key capabilities exposed as MCP tools:
- Station and exit lookup with wildcard search and flexible exit code matching
- Geospatial proximity search by coordinates or named landmark, powered by OpenStreetMap geocoding
- Radius-based exit discovery with in-memory Haversine distance ranking
- Transit line enumeration; all exits grouped by station across 9 MRT/LRT lines
- Domain-specific tools for accessibility, retail site analysis, logistics planning, emergency response, and tourist guidance

SecureChat
Built an open-source, production-ready starter template for real-time encrypted group chat — designed to be forked, deployed in one click on Replit, and extended without needing to rebuild the security layer from scratch.
Users authenticate via Replit OIDC (PKCE flow, no credentials required on the platform), then create public or private chat rooms, invite others via expiring single-use links, and exchange messages that are encrypted end-to-end in the browser before they ever leave the device. The server stores and forwards ciphertext only — it never has access to message plaintext at any point in the pipeline. Room keys are generated per-room, delivered to clients via an ECDH P-256 key exchange, and stored at rest wrapped with AES-256-GCM using a PBKDF2-SHA512 derived key (600,000 iterations). The server also relays WebRTC signaling — SDP offers, answers, and ICE candidates — as a blind relay for peer-to-peer connections.
The stack is a pnpm monorepo: vanilla HTML/CSS/JS frontend with no framework; Express 4 + Socket.IO 4 backend; PostgreSQL for message history and session persistence via connect-pg-simple; Zod for input validation; Pino for structured JSON logging. Authentication uses openid-client with PKCE.
Because all messages are encrypted before transmission, server-side content moderation is architecturally impossible. Instead, a shadow-ban profanity filter runs in the sender's browser before encryption — flagged messages are echoed locally to the sender but never transmitted to the server or other clients.
Security hardening includes a strict Content Security Policy, HSTS in production, Permissions-Policy, per-operation rate

limiting (joins, sends, history, invite creation and use), and input size caps on all socket payloads.
Fully open-sourced under MIT at github.com/leejaew/template-secure-chat and live at securechat-rt.replit.app

VisitKorea MCP
Built a production-ready Model Context Protocol (MCP) server that bridges AI agents with the Korea Tourism Organization's (KTO) official English tourism dataset, published on Korea's Public Data Portal (data.go.kr), enabling structured agent-driven access to tourism content across all 17 provinces of South Korea.
The server exposes a standardised set of tools that any MCP-compatible AI client (such as Claude, Manus AI, Cursor, or custom agents) can call to answer real-world questions about Korean tourism, without manual API integration required on the agent side.
Key capabilities exposed as MCP tools:
- Area-based and proximity search across 7 tourism categories: attractions, cultural facilities, restaurants, accommodation, shopping, leisure, and festivals
- GPS radius search using WGS84 coordinates with distance-ranked results
- Full-text English keyword search across all content types
- Festival and event discovery by date range and region
- Venue deep-dive tools returning opening hours, admission fees, rest days, parking, room types, menu items, and full image galleries
- Dual geographic code system support: new-system legal district codes and legacy area codes
- Delta sync feed for incremental cache refresh since a given modification date
The server is built on a two-tier architecture consisting of a Python/Starlette MCP layer handling protocol and upstream API calls, fronted by a Node.js/Express reverse proxy applying security headers, per-IP rate limiting, and structured logging. It is deployed on Replit as an open-source MIT-licensed project.

VisitKorea Medical Tourism MCP
Built a production-ready Model Context Protocol (MCP) server that bridges AI agents with South Korea's official Medical Tourism Open API (MdclTursmService), curated by the Korea Tourism Organization (KTO) and published on Korea's Public Data Portal (data.go.kr).
The server exposes 8 structured tools that any MCP-compatible AI client (such as Claude, Manus AI, Cursor, or custom agents) can call to discover, search, and retrieve detailed records for KTO-certified medical tourism facilities across South Korea, without any manual API integration required on the agent side. Built on FastMCP with Streamable HTTP transport, the server handles production concerns directly: async HTTP with connection pooling via httpx, a TTL response cache that delivers sub-5ms cache hits, a token-bucket rate limiter with fast-fail behavior to prevent upstream quota exhaustion, and strict parameter validation before any upstream request is made.
Key capabilities exposed as MCP tools:
- Region-based facility search by province, city, or legal district code
- GPS proximity search within a configurable radius (up to 20 km) using WGS84 coordinates
- Full-text keyword search across all facilities and regions
- Full dataset sync list for downstream database and ETL workflows
- Structured detail retrieval for any facility: common info, operational details, and medical-specific data including specialties, languages served, and reservation status
- Multilingual support across 5 languages: English, Japanese, Simplified Chinese, Korean, and Russian

VisitKorea Wellness Tourism MCP
Built a production-ready Model Context Protocol (MCP) server that connects AI agents to South Korea's official Wellness Tourism dataset, published by the Korea Tourism Organization (KTO) via the data.go.kr public data portal, covering spas, hot springs, jjimjilbang, traditional medicine clinics, healing meditation centres, beauty spas, and nature therapy sites across all regions of South Korea.
The server exposes a standardized set of tools that any MCP-compatible AI client (such as Claude, Manus AI, Cursor, or custom agents) can call to answer real-world questions about Korean wellness destinations, without manual API integration required on the agent side.
Key capabilities exposed as MCP tools:
- Region-based and GPS proximity search across all wellness venues nationwide, with distance sorting and a 20 km radius limit
- Full-text keyword search in Korean and 8 additional languages (English, Japanese, Simplified/Traditional Chinese, German, French, Spanish, Russian)
- Wellness theme filtering across 7 categories including hot spring/spa, jjimjilbang, Korean traditional medicine, and nature healing
- Administrative district code lookup for province and city/district-level filtering
- Full dataset sync list retrieval for building and maintaining local copies of the KTO dataset
- Venue detail tools returning operating hours, fees, facilities, accessibility info, homepage, and full photo galleries
Transport layer built on MCP Streamable HTTP in JSON response and stateless mode, making it compatible with gateway-style AI clients that enumerate tools at setup time and execute them later. Rate limiting, security headers, API key log redaction, and a server-side TTL cache are implemented to harden the production deployment.

MCP services connecting agents to structured public datasets, Public Data MCP Services

Built MCP services connecting AI agents to structured transit, tourism, medical tourism, wellness, and geospatial datasets. Includes multilingual search, proximity queries, caching, validation, and rate limiting.

Commercialization and use case development for enterprise AI platform, Allumni.ai Applied AI Agents
Led commercialization of MBLM's Generative AI platform into client-specific workflows. Designed prompt and persona systems, converted business workflows into AI agent patterns, built demonstrations, and supported enterprise onboarding.

Certifications

Master Builder, Replit — Jun 2026

Building with the Claude API, Anthropic — May 2026

Winner of Judges Favorite, Replit — Mar 2026

Prompt Design in Vertex AI, Google — Sep 2025

Odoo 18, Odoo — Jul 2025

ClickUp Expert, ClickUp — Jan 2024

Python Programming Beginner's Cookbook, — Dec 2023

Baseball Inspired Python Tutorial For Beginners, — Oct 2023

2nd Runner-Up for the Best Business Experience - airberlin #flighttoinnovation, airberlin — Dec 2016

Winner of Hack 4 Austin - Technology for the Good of Austin Kids, Hack 4 Austin — May 2014

Noticeable IT Business Model Award, The Entrepreneur Festival Korea — Jun 2010

Winner of the Official Development Assistance Business Competition, The Korea International Cooperation Agency (KOICA) — Nov 2008

USA Taekwondo State Championship Gold Medalist, United States Taekwondo Union — Jan 1993

USA Taekwondo State Championship Silver Medalist, United States Taekwondo Union — Jan 1992

Winner of Code for the Kingdom Hackathon, Oct 2013

Replit Agent,

Responsible Gaming Essentials, Front Office Sports

Sports Marketing Essentials, Front Office Sports

SABR Analytics Certification Level One, Society for American Baseball Research (SABR)

Professional Scrum Developer, Scrum.org

Professional Scrum Master II, Scrum.org

Professional Scrum Master I, Scrum.org

Titanium Certified Developer, Axway Developer

Titanium Certified Application Developer, Axway Developer

Forbes Technology Council Member and Contributor,

Skills

Agentic AI Development · Model Context Protocol (MCP) · Retrieval-Augmented Generation (RAG) · Large Language Models (LLM) · Natural Language Processing (NLP) · Application Programming Interfaces (API) · AI Product Development · Prompt Engineering · Artificial Intelligence (AI) · Hugging Face AI · Anthropic Claude · Open AI · Custom GPTs · Google Gemini · Physical AI · Workflow Automation · Open-Source Software · Enterprise Resource Planning (ERP) · Content Management Systems (CMS) · Customer Relationship Management (CRM) · Web Services API · eBay API · Facebook API · Geocoding · Foursquare · Google Maps API · Twitter API · LinkedIn API · iOS Development · iPhone Application Development · iPad Development · Full-Stack Development · FastAPI · Open API · PostgreSQL · React.js · HTML · Cascading Style Sheets (CSS) · JavaScript · PHP · Linux · Apache · Salesforce.com · Python (Programming Language) · Amazon Web Services (AWS) · Google Cloud Platform (GCP) · REST API · Node.js · MongoDB · NoSQL · Management · Public Speaking · Mentoring · Applied AI · Generative AI · Agent Orchestration · AI Evaluation