

Summary

AI application engineer experienced in building and deploying business-facing Agent workflows at Shopee Singapore. Combines analytics SQL, product requirement translation, and full-stack engineering, with 0-to-1 ownership across architecture, implementation, evaluation, and deployment. Strengths include translating business workflows into traceable AI applications and aligning requirements and delivery boundaries across multi-country BI teams.

Singapore-based | Employment Pass sponsorship required

Education

Nanyang Technological University | QS 2027: 12th globally | Nov 2025 - Jan 2027 (Expected) | Singapore
Master of Computing in Applied Artificial Intelligence, College of Computing and Data Science | No. 1 globally in Artificial Intelligence (2025 ShanghaiRanking)

Suzhou University | Sep 2019 - Jun 2023 | China
B.Eng. in Data Science and Big Data Technology; National Scholarship (top 0.2% nationwide)

Experience

 **Shopee Singapore** | Mar 2026 - Sep 2026
AI Data Analyst Intern, Buyer & Marketing Business Intelligence | Singapore

Analysis Agent Copilot · Application-layer owner · 0-to-1 Search / Compare / Draft

- Context and impact: Built for a data-request workflow spanning 9 analysts and 28 product managers, where historical requests, business SQL, and data topics were scattered and PMs relied on analysts to translate questions into structured requests; deployed into analysts' daily handling of real requests and built a Git-backed Analysis Knowledge Base with 146 source-, version-, and review-bound Case Bundles for traceable reuse of metric definitions and SQL.
- Ownership: Independently drove requirements, architecture, and delivery iterations; implemented the Svelte frontend, layered FastAPI backend (routes / services / repositories / renderers), Python ingestion pipeline, and Git-versioned knowledge storage.
- Agent architecture: Implemented allowlisted Search / Compare / Draft handlers for a constrained request-reuse workflow. The LLM handles intent and candidate explanations, while backend state controls retrieval, human selection, draft authorization, and evidence versions.
- Reliability and evaluation: Bound evidence to Git commits and content hashes, surfaced Changed / Preserved / Missing states before human review, and defined 20 exploratory regression tasks covering routing, retrieval, citation, parameter retention, safety, latency, and cost.

A/B Test Agent Skills · Internal Agent platform; used in business analysis

- Background: BI teams across countries repeatedly performed similar A/B Test analyses manually, reimplementing the same SQL and metric definitions without shared templates or acceptance criteria.
- Skill engineering and outcome: Helped define the Data Agent build SOP and package validated business SQL, experiment parameters, metric definitions, and acceptance criteria into constrained, reusable Skills used in real business analysis.
- Delivery and validation: Standardized date, region, experiment ID, and experiment-group ID; aligned scenarios and delivery boundaries with multi-country BI teams, authored the User Guide, and reconciled Skill outputs with baseline results under identical inputs. Preserved SQL Query Traces, applied bounded retry, and stopped at the defined error-policy boundary.

Jinken Vocational and Technical College	Mar 2023 - Oct 2025
Lecturer & AI Technology Lead	China
- Curriculum and practical delivery: Designed and delivered Python, C++, ROS Robotics Programming, and Introduction to Artificial Intelligence courses for 400+ students; built supporting practical modules and led hands-on ROS robotics training. - Engineering mentorship: Supervised capstone projects and work-placement internships for 23 students; led a ROS indoor autonomous-flight and obstacle-avoidance project through solution decomposition, system integration, issue diagnosis, and competition iteration, earning Third Prize at the 2024 RAICOM National Finals.	

Projects

BaZi Analysis Web App (Reliable LLM Workflow)	Jun 2026 - Present
Next.js 14, TypeScript, Zod, Redis, OpenAI-compatible API, NDJSON, Vercel	clinical-bazi-web.vercel.app
- Productization: Migrated a local domain-analysis Skill into a deployable full-stack Web application spanning automatic or manual intake, structured analysis, initial-review feedback, follow-up questions, and administrator audit; completed Vercel production deployment and acceptance testing across predefined end-to-end paths. - Reliability engineering: Separated a deterministic fact layer from LLM generation: the server computes the chart, binds evidence, and freezes facts, while the model generates candidate interpretations only. Zod contracts, semantic validation, a single controlled repair attempt, and fail-closed gates block unsupported or non-persistable output. - Long-running workflow: Implemented section-level NDJSON streaming, progress and heartbeat events, timeout cancellation, and terminal-state persistence; routed models by analysis stage, enabled fallback only after output-contract failure, and recorded the actual model, call count, and token usage. - Security and verification: Added signed sessions, rolling quotas, switchable local-file or Redis persistence, and administrator quality tracking; passed 112 deterministic and integration checks, TypeScript validation, and 23 build-time privacy-trace checks.	

ContextCue (Explainable Agent Memory and Continuity Recovery Prototype)	Jun 2026 - Jul 2026
TypeScript, Next.js, React	Live demo • GitHub repository
- Problem and delivery: Addressed cross-person memory leakage, unauthorized use of sensitive data, recall of expired or withdrawn content, and repeated context reconstruction after interruption; independently delivered the accessible Web prototype from problem definition and data modeling through interaction design. - Deterministic memory governance: Prototyped MemoryGate to structure person scope, provenance, confirmation status, privacy level, TTL, stance, and belief state; deterministic rules exclude wrong-person, expired, pending, external-opinion, or rejected records before generation. - Continuity and explainability: Built ContinuityEngine to reconstruct goals, settled decisions, rejected directions, open questions, and next steps from pre-interruption fragments; returned used and excluded evidence IDs with exclusion reasons for traceability. - Evaluation boundary: Designed 16 synthetic static scenarios covering person isolation, privacy and raw-content storage, TTL, evidence association, and tone safety; the prototype uses local deterministic logic and browser storage, not a production LLM evaluation.	

Technical Skills

AI Applications:	LLM Application Development, Agent Tool Use / Memory, RAG, Agent Harness, Eval Harness, Guardrails, Human-in-the-loop, Claude Code
Data & Analytics:	SQL, Presto / Trino, Hive, PySpark, pandas
Engineering:	Python, TypeScript, Svelte, FastAPI, Node.js, React / Next.js, Git / GitLab, Linux, CI
Languages:	Chinese (native); English (professional working proficiency; collaborated with multi-country BI teams in Singapore)
Certifications & Honors:	Software Designer; Huawei ICT Competition Global Final, Cloud Track, Third Prize (top 5%, Jun 2022)