

KENT (SUHAIL SHAH)

Senior Applied AI Engineer

Portfolio: kentdoki.dev · GitHub: github.com/kendoki · kentdoky@gmail.com · [LinkedIn](#) · fourpoket.com

SUMMARY

Applied AI Engineer with 8+ years of software engineering experience building production LLM applications, retrieval systems, and AI-powered products. Experienced in designing and shipping reliable AI systems across the full product lifecycle, including backend infrastructure, retrieval architectures, evaluation frameworks, observability, and deployment.

SKILLS

Applied AI: LLMs, Generative AI, AI Agents, RAG, Hybrid Search, Cross-Encoder Reranking, Embeddings, Vector Search, pgvector, Tool Calling, Structured Outputs, Prompt Engineering, LLM Evaluation (Evals), Semantic Caching, Model Routing, Azure OpenAI

Backend & APIs: Python, FastAPI, Pydantic, TypeScript, Node.js, PostgreSQL, Redis, SQLAlchemy, Celery

Data & Search: PostgreSQL Full-Text Search, PostGIS, Row-Level Security, OCR Pipelines, Document Processing

Cloud & Observability: Docker, AKS, GitHub Actions, OpenTelemetry, Datadog, CI/CD

EXPERIENCE

Senior Applied AI Engineer · Etiqa Insurance

Apr 2024 - Nov 2025

- Reduced policy lookup time by over 90% (10–20 minutes to under a minute) by building a citation-grounded enterprise AI knowledge platform for claims operations with hybrid retrieval and structured lookups.
- Implemented hybrid retrieval, cross-encoder reranking, and structured lookups for enterprise knowledge search.
- Architected and developed document ingestion and intelligence pipelines using Python, FastAPI, OCR, and pgvector.
- Established evaluation frameworks, observability, and citation-grounded AI responses.
- Designed and operated production infrastructure on AKS using Docker, GitHub Actions, and Terraform.

Senior Applied AI Engineer · Freelance

Apr 2023 - Mar 2024

- Built a multi-tenant AI knowledge platform integrating enterprise knowledge sources for citation-grounded conversational search.
- Integrated Notion, Google Drive, Confluence, websites, and document repositories into a unified retrieval pipeline.
- Improved internal retrieval precision from ~60% to the low-80% range by implementing hybrid search, cross-encoder reranking, and evaluation-driven optimization.
- Built evaluation frameworks, analytics, and observability to support production AI reliability and cost optimization.

Senior Software Engineer · Accenture

Nov 2020 - Mar 2023

- Delivered enterprise workflow automation and document intelligence platforms using TypeScript and Python.
- Led development of configurable workflow systems adopted across multiple business units.
- Improved operational efficiency through automation, OCR/NLP processing, and streamlined review workflows.

Software Engineer · Sunway

Nov 2019 - Oct 2020

- Built backend APIs, operational dashboards, and real-time monitoring systems.
- Developed scheduling and workforce management solutions using SQL and PostGIS.
- Reduced manual planning effort from hours to minutes through automation.

Software Engineer · Vistra

Aug 2017 - Aug 2019

- Developed and maintained full-stack JavaScript web applications for business operations.

PROJECTS

fourpoket · AI coding agent · live product

Jan 2026 - Apr 2026

fourpoket.com · [npm: fourpoket](https://npm.com/fourpoket)

- Designed and shipped a production AI coding agent CLI with multi-provider orchestration, AST-based code editing, retry/fallback strategies, and per-call cost tracking, over a three-phase pipeline on an AST engine with targeted reads and surgical writes.
- Built and launched the end-to-end SaaS platform, including authentication, Stripe billing, credit enforcement, admin tooling, and production monitoring.

EDUCATION

Bachelor of Information Technology (Honours) · Software Systems Development (RSD)

2017