

NECMI KURAL

Aotearoa (New Zealand) | linkedin.com/in/necmikural | github.com/NecmiKural | necmikural.github.io

PROFILE

Software engineer. I build tool-calling agents and retrieval systems, and measure them with evaluations that replay from recorded runs. For over two years I have worked remotely as a software developer for a fintech company, owning a partner-management platform end to end: C#/ASP.NET Core microservices, PostgreSQL, a Vue 3 + TypeScript dashboard and Flutter apps, shipped with Docker and CI/CD. Postgraduate Diploma in Computer Science with Master of AI coursework, University of Waikato, 2026.

CORE SKILLS

LLM systems: Tool-calling agents with strict-schema tools in a hand-rolled loop with step, token and cost limits; a read-only MCP server; RAG with PostgreSQL and pgvector over dated sources, with citations and figures checked in code; prompt-injection tests, defence ablations and personal-data redaction.

Evaluation: Repeated live evals across models, frozen eval sets with held-back questions, a pre-registered experiment with its failed tests published, CI replay gates, and README numbers generated from recorded results.

Backend and data: C#, ASP.NET Core, REST, Entity Framework Core, Dapper, PostgreSQL; Python (pandas, NumPy, scikit-learn, PyTorch).

Front end and mobile: TypeScript, Vue 3, React, Next.js; Flutter (Dart).

Delivery: Docker, GitHub Actions, Linux, nginx, Git, xUnit, code review; coding agents in daily use with repository context files (CLAUDE.md, AGENTS.md).

SELECTED PROJECTS

Bank Reconciliation Agent | Personal project, v1.1.0 | Python, PostgreSQL, pgvector, MCP, Docker

A rules-first tool-calling agent that matches small-business bank lines to invoices on synthetic data. Code verifies every proposal, the model cannot write to the ledger, and a person approves before anything posts. With deepseek-flash: 0 wrong matches in 460 agent match proposals (582 agent decisions) over 3 live runs on each of two synthetic businesses. 0 of 40 injected lines that ask for a different or an extra invoice succeeded in any run.

github.com/NecmiKural/bank-reconciliation-agent

Provenance-Tiered RAG | Personal project, v0.3.0 | Python, PostgreSQL, pgvector

Question answering over New Zealand GST law, German VAT law (UStG and UStDV) and Immigration New Zealand's Operational Manual, with dated sources in tiers: statute, official guidance and untrusted notes. Code checks every citation and figure against passages in force on the date asked about; immigration answers are information from published instructions, not immigration advice. With deepseek-flash: 96 of 96 answers to answerable questions were correct (95% lower bound 96.2%) over 3 runs of the development questions; held-back NZ GST questions, answerable correct: 27 of 30 (90.0%).

github.com/NecmiKural/provenance-tiered-rag

mcp-code-mode | Personal project, v0.1.0 | Python, MCP, Docker

Runs model-written Python against any MCP server's tools in a locked-down Docker sandbox; every call passes a host-side bridge that checks the allowlist, the tool's JSON Schema, argument size and a call budget. Where it was measured, in bank-reconciliation-agent: containment held in every run and every probe, and code mode was cheaper on average, but the pre-registered efficiency hypothesis failed for both models tested.

github.com/NecmiKural/mcp-code-mode

crossfade | Personal project, v0.1.0 | TypeScript, Next.js, PostgreSQL, Envoy

Canary releases with automatic rollback for plain Docker hosts: Envoy moves traffic to a candidate in steps and rolls back on health, convergence or error-budget breaches, driven by a small control plane. Compose end-to-end test in CI; it has not been audited.

github.com/NecmiKural/crossfade

MACHINE LEARNING COURSEWORK, UNIVERSITY OF WAIKATO

- **BirdCLEF 2026 audio classification** (Deep Learning, COMPX525, individual): Tests whether AudioLDM2 synthetic clips for rare classes help two classifiers; they helped BirdNET embeddings with a linear classifier and hurt a CNN trained from scratch. BirdNET embeddings + ML reach 0.8940 validation macro ROC-AUC with synthetic enrichment and 0.8869 without, on one fixed 80/20 split of train_audio with one seed; this is not a Kaggle leaderboard score.
- **Cell-culture image classification** (Deep Learning, COMPX525, individual): Heavily imbalanced microscopy classification comparing a custom CNN, a ViT trained from scratch, fine-tuned pretrained models and embedding classifiers. Best clean result: 76.88% validation accuracy for a single fine-tuned ConvNeXt-Tiny on a 15% stratified holdout; 5-fold out-of-fold accuracy 73.04%.

- **Streaming ensemble under a memory budget** (Machine Learning for Data Streams, COMPX523, individual): An online ensemble for a drifting, imbalanced stream under a hard 250 KB memory budget, with a post-mortem of the memory-measurement bugs that put the submitted version over the limit, and a re-run with the cap enforced.
- **kNN for data streams** (Machine Learning for Data Streams, COMPX523, individual): Online kNN with three window strategies; on the imbalanced, drifting RBF stream, plain kNN with a small window collapsed to ~66%, while per-class windows held at ~85% (single runs per setting).
- **Transit peak-hour drift detection** (Machine Learning for Data Streams, COMPX523, group project of 6): My part, Role D: the adaptive boundary detector, ADWIN and Page-Hinkley drift tests and schedule alignment. Data: three months (March to May 2024) of historical 5-minute estimated boarding counts for 10 Salvador bus stops (SUNT, CC BY 4.0), replayed as a stream.

OPEN SOURCE

- m365-copilot-proxy (upstream contributions, TypeScript): three merged pull requests: Windows agent fixes, pinning each conversation to one backend, and image input.

EXPERIENCE

Software Developer (remote), fintech company 2024 to present

- Own a partner-management platform end to end: C#/ASP.NET Core microservices behind an API gateway, PostgreSQL through Entity Framework Core and Dapper, scheduled background jobs and a Vue 3 + TypeScript dashboard, deployed with Docker and nginx on Linux.
- Integrated third-party trading platform APIs into commission and reporting pipelines, with reconciliation that surfaces records that do not match between systems.
- Sole developer of the company's Flutter apps for iOS and Android; xUnit tests, CI/CD and code review in agile sprints; coding agents in daily use with repository context files.

Mobile Developer, brokerage firm, Istanbul 2022 to 2023

- Built a Flutter crypto-exchange app with backend API integration; set up Git workflows, Dockerised the applications and deployed them on Google Cloud.

Web Developer and Team Lead, web agency, Istanbul 2021 to 2022

- Led a small web-development team; built e-commerce sites and Laravel backend APIs.

EDUCATION

- Postgraduate Diploma in Computer Science with Master of AI coursework, University of Waikato, 2026
- Bachelor of Science in Chemistry, Istanbul University, 2024

OTHER

- Chairperson of the Board, IAESTE Cerrahpasa, Istanbul University (2023). Languages: English (professional), Turkish (native).