Miles Thompson

Machine Learning & AI · Data Sovereignty & Responsible AI · Tech Leadership

Wellington, New Zealand (happy to travel) · +64 21 539 843 · utunga@gmail.com · LinkedIn · GitHub

SUMMARY

Mathematician turned ML engineer and technology leader. For six years at Te Hiku Media I built foundational speech language models for Māori and English while demonstrating Indigenous data sovereignty. Our small AI Research Lab is known worldwide. Before that, I was the founding CTO of CreditSights (acquired by Fitch Ratings for $220m) and Contract CTO of AI start-up Chatterize (acquired 2022). I have presented at NeurIPS, ICML and AAAI and hold a US patent. I thrive on hard problems and realising the vision of people smarter than me.

SKILLS

- **Responsible AI:** evaluation and benchmark design, runtime agent governance and drift detection, data isolation for AI agents
- **Research:** experiment design, baselines and ablations, Bayesian modelling (Stan), statistics, technical writing and conference presentation
- **Machine learning:** PyTorch, NeMo, scikit-learn, TensorFlow; speech recognition and synthesis, Transformers / Conformers, NLP, LLM agents
- **Programming / development:** Python, C#, SQL, TypeScript; full-stack web development, back-end services and APIs
- **Technical leadership:** building teams from scratch, hiring and mentoring, technology strategy and budgets, working with researchers and community partners
- **Cloud ops / MLOps:** Kubernetes, Kubeflow, OpenStack (Catalyst Cloud), AWS; CI/CD and production ML

EXPERIENCE

Agentic Systems Engineer | Digicryptids (AI venture studio) Oct 2026 – present

- Designed oven-mitt, an open-source vault that lets AI agents work with data without it ever entering the model's context, turning the Kaitiakitanga Licence's data-protection clause into an enforced technical control.

Contract CTO, Agentic AI Governance | ChurPrimo Mar 2026 – Oct 2026

- Built a runtime-governance system that reads LLM agents' reasoning traces, detects drift and corrects it mid-task. Presented it at the AAAI Spring Symposium 2026 (Multi-Agent Systems track).

Senior Machine Learning Engineer / Team Lead | Te Hiku Media (Papa Reo) Apr 2020 – Jul 2026

Iwi-led AI lab building speech and language technology for te reo Māori, NZ English and Pacific languages under Indigenous data sovereignty. Featured in Karen Hao's "Empire of AI"; CEO named in the TIME100 Most Influential People in AI (2024).

- **Data sovereignty:** Trained models only on data for which kaitiakitanga (responsible governance) of data concerns were correct. Moved all training and inference onshore to Catalyst Cloud and on-prem so the data stayed under iwi (tribal) control. Presented the lab's work at ICML 2023.
- **Firsts:** As lead AI developer, pioneered the world's first Māori speech synthesis model and pretrained bilingual Māori / NZ English foundation speech models from scratch.
- **Interpretability:** Conceived a pronunciation-feedback model that works by introspecting the internal activations of a speech recognition RNN (NeurIPS 2021 poster). Shipped it in the Rongo app, #1 in Education on the NZ App Store.
- **Data & production:** Built pipelines combining human judgement with trained automatic quality measures over thousands of hours of curated audio, plus the CI/CD and benchmark monitoring that put models into production.
- **Leadership:** Built the lab's data science team from one engineer, hiring and mentoring researchers and engineers.
- **Evaluation:** Contributed to the research behind "Fit for our purpose, not yours" (NeurIPS 2024), a benchmark for a low-resource Indigenous language. On held-out, purpose-built test sets our models beat every commercial vendor we compared, including on NZ English.

Contract CTO | Chatterize (later Speakia, acquired 2022) May 2019 – Apr 2020

- Took an AI language tutor for children from idea to launch in China, years before AI tutors went mainstream. Recruited the China-based team; growth reached thousands of sign-ups a week and a seed round of more than $1.5m.

Machine Learning Engineer (contract) | Te Hiku Media Mar 2018 – Jun 2019

- Pioneered the world's first end-to-end speech recognition system for te reo Māori, the proof of concept that helped us win a NZ$13m MBIE research grant.

Artist-Programmer / ML Engineer (NLP) | The Great Outdoors (Annie Dorsen) Sep 2016 – Sep 2017

- Built the text engine for Annie Dorsen's algorithmic theatre piece: it harvests recent internet comments, scores them for entropy and semantic fit (word2vec, LSTM) and writes a unique script for every performance. The show toured, opening in NYC, Paris and Amsterdam.

Founding CTO, then Senior Technical Architect | CreditSights Inc, New York Nov 2002 – 2018

First technology hire at a start-up that grew into the leading independent voice in financial research: ~12 to ~200 people, $50m+ revenue, acquired by Fitch Ratings for $220m (2020).

- Built the technology department from scratch and partnered with quants and analysts to turn their models into products, letting a small independent compete with far better-funded "sell-side" broker-dealer research.

Selected earlier work 2000 – 2020

- Mutual Aid World (2020): co-founded a COVID-response project and coordinated 400+ volunteer developers to deliver food in Los Angeles.
- Reserve Bank of New Zealand (2019): more than 5× speed-up of the Bank's financial-returns system, ahead of schedule and under budget.
- Faction / W4GL (2018–19): co-authored a white paper on resilient distributed networks, with Bayesian models (Stan) of how reliable information propagates in a crisis.
- Groundfacts (2011–12): founded a team to crowdsource and filter reliable crisis information from social media after the Christchurch earthquake, working with university crisis-management researchers.
- Hashmapd (2010–12): co-founded an early deep learning start-up with Drs Abraham and Frean (autoencoder semantic hashing, t-SNE maps).
- Medigenesis (2000–08): architected a medical collaborative-filtering system recommending treatments that worked for "others like you"; US patent co-inventor.

PUBLICATIONS & TALKS

- NeurIPS 2024: "Fit for our purpose, not yours: Benchmark for a low-resource, Indigenous language" (research contributor)
- ICML 2023: "Enabling a sovereign digital future for Indigenous languages" (presenter)
- NeurIPS 2021 poster: "Scoring pronunciation accuracy via close introspection of a speech recognition recurrent neural network" (co-author)
- AAAI Spring Symposium 2026, Multi-Agent Systems track: runtime agent governance demo (presenter)
- US patent 7,676,384: Medical collaborative filtering and cluster analysis system (co-inventor)

EDUCATION

BSc (Hons I), Mathematics, University of Otago. First-class honours. Honours project: "A Neural Statistical Hybrid".
BA, Sociology, University of Canterbury.