Kaitiakitanga in Machine Learning:
A Budget-Aware Approach to Fairness in Bilingual RAG

Executive Summary

This report examines fairness in AI systems serving both English and Te Reo M■ori. Through a budget-aware Retrieval-Augmented Generation (RAG) system, this research explores how resource allocation strategies can reduce performance disparities between high-resource (English) and low-resource (M■ori) languages. The project reveals that improving retrieval quality through semantic embeddings yields greater fairness gains than budget allocation mechanisms alone. This work contributes to conversations about M■ori data sovereignty, ethical AI design, and language justice.

1. Motivation: Language Justice and Kaitiakitanga

Natural language processing systems exhibit systematic performance disparities, with low-resource languages like Te Reo M■ori (185,000 speakers) consistently underperforming compared to English. Contemporary AI systems, trained predominantly on English data, risk perpetuating "algorithmic colonialism"—the extension of historical power imbalances into computational infrastructure (Couldry & Mejias, 2019).

The concept of **kaitiakitanga**—guardianship and stewardship—provides a culturally grounded framework for this technical challenge. In tikanga M■ori, kaitiakitanga encompasses responsibility to protect taonga (treasures), which explicitly includes te reo M■ori itself (Waitangi Tribunal, 2011). This project adopts kaitiakitanga as both metaphor and method: just as traditional kaitiaki actively manage resources to ensure equitable access, a budget-aware RAG system allocates computational resources to protect M■ori language interactions with AI. This embeds cultural values at the architectural level rather than treating fairness as post-hoc optimization.

Research Questions: (1) Can dynamic budget allocation reduce performance gaps between English and M■ori queries? (2) What retrieval strategies most effectively support fairness? (3) How do fairness interventions align with indigenous data sovereignty principles?

2. Methods: Three Fairness Philosophies

The kaitiaki-planner implements a RAG pipeline with query processing, budget planning, semantic retrieval (paraphrase-multilingual-mpnet-base-v2 with keyword boosting), and answer generation via Claude (Anthropic's LLM). Three budget allocation strategies operationalize distinct fairness theories:

Condition 1: Uniform (Equality). Allocation: top_k=5 for all queries. Philosophy: Formal equality—identical resources regardless of need. Represents most current RAG systems.

Condition 2: Language-Aware (Equity). Allocation: top_k=8 for M■ori, top_k=5 for English. Philosophy: Equity based on resource scarcity—additional support for structurally disadvantaged groups. Acknowledges M■ori queries face retrieval challenges from smaller training data and fewer native embeddings.

Condition 3: Fairness-Aware (Intersectional Equity). Allocation: top_k=8 for M■ori OR complex queries. Philosophy: Intersectional fairness recognizing multiple axes of disadvantage (language, task difficulty). Operationalizes **manaakitanga** (care, support) by providing differentiated support based on structural barriers.

Evaluation: 30 labeled queries (15 English, 15 M■ori) covering New Zealand topics. Metric: Grounded Correctness (GC)—binary measure of citation accuracy. Baseline: BM25 retrieval achieved 100% for English but only 46.7% for M■ori (53.3 point gap). Improved system replaced BM25 with semantic embeddings plus 30% keyword boost.

3. Results: Quality Trumps Allocation

Performance Improvement: English maintained 100% accuracy; M■ori improved from 46.7% to 80.0% (+71% relative improvement). Performance gap reduced from 53.3% to 20.0% (62% reduction).

The Null Result: All three budget allocation strategies produced identical results—mean GC of 0.900 (27/30 queries) across all conditions. English: 15/15 (100%) in all conditions; M■ori: 12/15 (80%) in all conditions. Statistical tests confirmed

no difference (ANOVA F=0.000, p=1.0000).

Interpretation: Once semantic embeddings correctly ranked the gold-standard passage within top 5 results, increasing budget to top_k=8 provided no additional benefit. The correct passage was already visible; retrieving more documents introduced noise rather than signal. This reveals that **fairness-by-design** (choosing multilingual embeddings) proved more effective than **fairness-by-allocation** (dynamically adjusting top_k). Budget reallocation serves as crisis management; architectural quality is preventive care.

Persistent Gap: While reduced, a 20-point gap persists (t-test p=0.0719, Cohen's d=0.683). Three M■ori queries failed across all conditions, revealing limits requiring further investigation.

4. Ethical, Cultural, and Societal Implications

4.1 M■ori Data Sovereignty

This project operates within the framework articulated by Te Mana Raraunga (M■ori Data Sovereignty Network). Three principles inform this work:

Rangatiratanga (Authority/Self-Determination): M■ori should govern how te reo M■ori is represented in AI systems. True rangatiratanga requires M■ori communities not just as evaluation subjects but as co-designers, determining system priorities and acceptable trade-offs.

Whakapapa (Relationships/Contextualization): Data must be understood in relational context. The corpus pairs English and M■ori passages about shared referents (e.g., Tongariro), acknowledging these reflect different epistemologies—m■tauranga M■ori versus P■keh■ knowledge systems. RAG systems treating these as interchangeable risk flattening important distinctions about knowledge authority.

Kaitiakitanga (Guardianship): Beyond technical adequacy, genuine kaitiakitanga asks: Who benefits? Who is harmed? Current implementation improves M■ori performance to 80%, but still underperforms English. Is 80% acceptable? Under what theory of justice is systematic underperformance tolerable?

4.2 Ethical Philosophy: Distributive Justice Theories

The three experimental conditions operationalize distinct theories:

Utilitarian Equality (Uniform): Maximize average performance, treating all queries identically. Efficient but ignores structural inequalities, perpetuating existing advantage.

Rawlsian Equity (Language-Aware): Prioritize improvements for the least advantaged group (Rawls, 1971). Allocating additional budget to M■ori queries embodies the "maximin" principle—optimize for the worst-off position.

Capabilities Approach (Fairness-Aware): Drawing from Sen (1999), this approach asks not whether people have equal resources, but whether they have equal capability to achieve valuable outcomes. Identifies multiple barriers (language, complexity) constraining capability, then allocates resources to equalize effective capability rather than formal inputs.

The null result complicates this philosophical landscape: when all approaches yield identical outcomes, the choice becomes moot in practice. This reveals that **fairness interventions are context-dependent**. Sophisticated allocation mechanisms matter when resources are scarce and quality uneven; they become irrelevant when baseline quality meets user needs. The ethical imperative shifts from 'how do we allocate resources fairly?' to 'how do we ensure sufficient quality that allocation becomes unnecessary?'

4.3 Machine Consciousness and Epistemic Violence

While debates about machine consciousness typically focus on whether AI systems possess subjective experience (Chalmers, 1995), this project highlights a different dimension: the **ethics of representation** in systems processing language. Even if Claude lacks consciousness, it serves as an epistemological intermediary—shaping how users access knowledge.

When a RAG system fails to retrieve correct M■ori passages, it performs a kind of **epistemic violence**, reinforcing the false belief that information is less accessible in te reo M■ori. This matters particularly for younger generations: if digital systems consistently deliver better results in English, users face pragmatic pressure to abandon te reo, accelerating language loss.

The concept of **linguistic agency**—users' ability to interact with technology in their language of choice without penalty—becomes crucial. Fairness isn't just about evaluation metrics; it's about whether computational infrastructure supports or undermines indigenous language vitality. The 20-point gap represents not merely a technical deficiency but a social harm requiring urgent redress.

4.4 Cross-Cultural Implications

While grounded in Aotearoa's context, this work has implications for the estimated 7,000 languages worldwide, most "low-resource" from NLP's perspective (Joshi et al., 2020). Three lessons emerge:

1. Architecture Matters More Than Tweaking. Replacing BM25 with semantic embeddings yielded 71% improvement for M■ori retrieval—far more effective than any allocation strategy. For language communities considering AI interventions, prioritize foundational models respecting linguistic diversity over downstream optimizations within already-biased systems.

2. Measurement Shapes Outcomes. The Grounded Correctness metric deliberately evaluates citation accuracy, not fluency, reflecting understanding that retrieval quality precedes generation quality. Other evaluation contexts might prioritize cultural appropriateness or preservation of traditional metaphors. Fairness metrics must be co-designed with communities to reflect their priorities.

3. The "Good Enough" Trap. An 80% success rate for M■ori queries might seem acceptable—better than 46.7%. But "good enough" thinking entrenches inequality when one group receives 100% performance. The remaining 20-point gap represents structural unfairness: some users enjoy perfect service while others experience regular failures. Acceptability cannot be judged in isolation but only in comparison to best performance achieved.

4.5 Limits of Technical Solutions

This project illuminates the limits of technical interventions divorced from structural change. Even a perfectly fair RAG system operates within broader inequalities: most training data comes from English sources; most AI research papers are published in English; most companies hiring NLP engineers expect English-language credentials. Improving one system's M■ori performance cannot substitute for systemic efforts:

Data Justice: Investing in M■ori-language corpus development, oral history digitization, and indigenous-language web content. **Educational Justice:** Funding M■ori-language immersion education and bilingual STEM resources. **Economic Justice:** Supporting M■ori-led tech companies and procurement policies favoring bilingual AI systems.

The principle of **whanaungatanga** (kinship, relationship) reminds us that technology exists within social webs. A RAG system cannot save a language; only communities can, through intergenerational transmission, cultural revitalization, and political power. Technology can support or hinder these efforts, making algorithmic fairness deeply political.

5. Conclusion: Toward Language Justice in AI

This capstone demonstrates that meaningful fairness improvements in bilingual RAG systems require architectural investment in retrieval quality, not merely clever resource allocation. By improving M■ori query performance from 46.7% to 80% through semantic embeddings and keyword boosting, the system substantially reduced but did not eliminate the performance gap with English.

From a kaitiakitanga perspective, this suggests a clear mandate: those building AI systems for multilingual contexts must prioritize foundational model quality respecting linguistic diversity. Budget allocation strategies serve as contingency measures, useful when quality falls short but insufficient as primary fairness mechanisms.

The persistent 20-point gap, though statistically marginal ($p=0.0719$), remains practically significant. Future work must develop targeted interventions including query expansion, cross-lingual retrieval, and community evaluation by native speakers assessing cultural appropriateness beyond factual correctness.

This work contributes to an emerging paradigm of **indigenous AI ethics**—frameworks centering indigenous epistemologies, data sovereignty, and language vitality rather than treating these as afterthoughts. As AI systems become ubiquitous, the question of whose languages are supported becomes a question of whose knowledge endures. In Aotearoa and globally, the answer must be: all languages, equitably. Anything less perpetuates digital colonialism, using new technologies to repeat old injustices.

The mauri (life force) of te reo M■ori depends not on algorithms but on speakers—children learning, elders teaching, communities thriving. Yet in a digitally mediated world, computational kaitiaki have a responsibility: to ensure technology

supports rather than suffocates indigenous language vitality. This capstone represents one small effort in that direction, demonstrating both the possibilities and the profound work remaining.

References

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of ACL 2020*, 6282-6293.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Te Mana Raraunga. (2018). Principles of M■ori data sovereignty. Retrieved from https://www.temanararaunga.maori.nz/

Waitangi Tribunal. (2011). *Ko Aotearoa T■nei: A report into claims concerning New Zealand law and policy affecting M■ori culture and identity* (Wai 262). Wellington: Legislation Direct.

Author: *Alex* | ***Project:*** *Kaitiaki-Planner* | ***Date:*** *October 2025*