Proceedings of the 4th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2025.BJ30766

The Role of Financial Mathematics in Data Protection and Fairness in Technology: Based in a Case Study of Te Hiku Media

Yi Gu

The London School of Economics and Political Science(LSE), London, The United Kingdom
y.gu.sirius@gmail.com

Abstract. This analysis explores how Te Hiku Media, a Māori-led non-profit organization, integrates interdisciplinary strategies to protect and revitalize the Māori language while actively resisting the forces of data colonization driven by large technology corporations. Central to their approach is the development of in-house machine learning models that leverage semi-supervised learning methods, which significantly reduce the need for extensive labeled datasets. By adopting this strategy, Te Hiku maintains linguistic sovereignty, safeguards cultural heritage, and minimizes reliance on external corporate infrastructures. Beyond the technical dimension, the organization applies financial tools such as cost-benefit analysis, Bayesian probability, and multivariate regression to support risk-aware decision-making in areas of data ownership, potential commercialization, and the long-term preservation of cultural value. The initiative also foregrounds critical ethical considerations, including the avoidance of algorithmic bias, the prevention of price discrimination in educational access, and the enforcement of strict community-led governance structures to ensure data privacy and cultural accountability. Te Hiku Media's work demonstrates how localized, ethical applications of artificial intelligence can simultaneously promote cultural continuity and advance social equity. Their model provides a replicable pathway for other Indigenous communities navigating similar technological, cultural, and geopolitical challenges in the rapidly evolving digital age.

Keywords: Te Hiku Media, Financial Mathematics, Data Protection

1. Introduction

Te Hiku Media, a small non-profit radio station based in New Zealand, has become a beacon of hope for the revitalization of the Māori language. With Māori speakers dwindling and the cultural significance of the language at risk of being lost, Te Hiku Media has taken up the mantle of preserving this valuable cultural asset. The organization's efforts to safeguard the Māori language not only have social and cultural impacts but also intersect with broader discussions on data protection, technological fairness, and the financial implications of artificial intelligence (AI) and machine learning in today's rapidly evolving technological landscape.

DOI: 10.54254/2754-1169/2025.BJ30766

At the heart of the conversation is the notion of “data colonization”—a phenomenon where large tech corporations extract valuable data from Indigenous communities, monetize it, and fail to provide fair returns. This issue has become increasingly relevant as AI systems continue to rely on vast amounts of data, including language data. In many cases, the communities from which this data is taken see little to no benefit, while global corporations capitalize on it. Te Hiku Media’s efforts represent an attempt to break this cycle and ensure that the Māori people retain control over their language data.

Financial mathematics plays a critical role in this struggle. Through the application of economic models and risk assessments, financial mathematics can help Te Hiku Media weigh the benefits and potential risks associated with selling or sharing their language data. By using methods such as cost-benefit analysis, Bayesian probability, and multivariate regression models, Te Hiku Media can make informed decisions about how to manage their data and ensure that the economic benefits of their language resources remain within the community. Furthermore, by keeping the data within local control, Te Hiku Media avoids the commercialization of Māori language data for profit—an outcome that would have detrimental social, cultural, and financial consequences for the Māori people.

Financial mathematics provides essential tools for analyzing the economic risks and rewards associated with data ownership, especially when it comes to the commercialization of sensitive or cultural data. In the case of Te Hiku Media, the question of whether to sell Māori language data to multinational tech companies raises significant financial and ethical concerns. Without careful analysis, these decisions could result in the exploitation of Māori language resources, with profits flowing to global corporations rather than the Māori community.

Financial mathematics can aid in modeling the various risks and returns of such a decision. Cost-benefit analysis, a fundamental concept in financial mathematics, could be used to assess the financial trade-offs between retaining control of the language data and selling it for potential profits. However, the financial calculus goes beyond mere dollars and cents. Key variables, such as the long-term cultural value of the language, the potential for social inequities, and the risk of "data colonization," must be incorporated into the model.

One important tool in this context is Bayesian probability, which helps in assessing the uncertainty associated with the risks of commercializing Māori language data. By estimating the probability of various outcomes—such as the possibility that big tech companies will fairly compensate the Māori community for their data, or that the data will be exploited for profit—Te Hiku Media can make a more informed decision. Bayesian analysis allows for a dynamic understanding of the risk, where the probabilities are updated as new data becomes available.

Additionally, multivariate regression models could be used to analyze the relationship between different variables in this decision-making process. For example, the regression model could assess the impact of various factors—such as the level of data protection, the cultural significance of the language, or the financial gain from selling the data—on the overall economic benefits for the Māori community. This would allow Te Hiku Media to predict the potential outcomes of different courses of action and choose the one that maximizes both economic and cultural benefits.

2. Machine learning and semi-supervised learning: reducing big tech's influence

Machine learning (ML) has become an integral part of data processing and language preservation in recent years. Te Hiku Media’s use of ML models, especially in the context of preserving the Māori language, showcases a novel approach to counteracting the influence of big tech companies. By

DOI: 10.54254/2754-1169/2025.BJ30766

developing their own machine learning tools, Te Hiku Media is not only fostering the growth of Māori language resources but also ensuring that control over the data stays within the community.

One of the most significant advantages of Te Hiku Media's approach is the use of semi-supervised learning, a machine learning technique that requires less labeled data to train models effectively. Unlike traditional ML models, which rely on vast amounts of labeled data for training, semi-supervised learning allows the Te Hiku team to make use of a smaller dataset while still achieving high accuracy. This is particularly important for indigenous languages like Māori, where labeled datasets may be scarce due to historical neglect and the lack of technological resources.

Semi-supervised learning reduces the reliance on external data providers, particularly large corporations that might seek to exploit Māori language data for profit. By developing their own language models, Te Hiku Media is ensuring that the Māori community retains ownership of its language data and has the means to revitalize and promote its language independently of commercial pressures. This approach also minimizes the risks of data commodification, where private and culturally sensitive data is extracted, used for profit, and potentially lost to history.

Moreover, the use of machine learning, coupled with the application of financial mathematics to assess risks and outcomes, allows Te Hiku Media to make strategic decisions about how to best utilize the data they collect. The integration of machine learning and financial analysis ensures that Te Hiku Media can strike a balance between preserving the integrity of the Māori language and navigating the financial realities of operating in a tech-driven world.

3. Fairness and equity in data usage

The issue of fairness in data usage is central to the conversation about data protection and the commercialization of indigenous languages. As big tech companies continue to expand their use of AI and machine learning, there is growing concern about how data is extracted from marginalized communities and how those communities benefit—or fail to benefit—from the value created by this data.

In the case of Te Hiku Media, fairness and equity in data usage are crucial not only from a financial standpoint but also from a social justice perspective. The exploitation of Māori language data by large corporations could lead to increased inequality in access to language resources, particularly in educational contexts. For example, if the Māori language data were sold to third-party companies, there is a risk that the educational tools and resources developed using that data could be sold at high prices, excluding lower-income Māori families from accessing them.

Price discrimination, in which different groups are charged varying amounts based on their socio-economic status, is a major concern in the commercialization of indigenous data. If Māori families are unable to afford these educational tools, the language could become further marginalized. Conversely, if the Māori community retains control over their language data, they can ensure that language revitalization efforts are accessible to all, regardless of economic background.

Furthermore, by keeping language data under local control, Te Hiku Media can avoid potential biases that may arise in AI systems developed by external organizations. AI algorithms, particularly those based on data that are not representative of diverse communities, are often prone to biases that can perpetuate social inequalities. By developing their own language models, Te Hiku Media is not only ensuring that Māori language data is used fairly but also protecting against the risks of bias that could affect the language’s development in an increasingly AI-driven world.

DOI: 10.54254/2754-1169/2025.BJ30766

4. The broader ethical and societal impacts of AI and data protection

The rise of AI and its ability to process vast amounts of data has raised fundamental ethical questions regarding privacy, fairness, and the societal impact of these technologies. AI systems have the potential to change the way society functions—especially in areas like education, healthcare, and even politics. However, as AI becomes more ubiquitous, it is essential to consider the ethical implications of using AI in sensitive areas, such as language preservation.

One of the primary concerns is the potential for data exploitation. The rapid growth of AI has led to concerns about how companies use, store, and share personal data. In recent years, we have seen numerous instances of data breaches, privacy violations, and monopolies on personal data, such as Facebook's control over user data for targeted advertising. These examples highlight the dangers of allowing large corporations to control sensitive data, particularly data that belongs to marginalized or indigenous communities.

For Te Hiku Media, the protection of Māori language data is not just a matter of financial gain—it is also a matter of cultural and social justice. The exploitation of indigenous data for profit by large corporations can have long-lasting effects on the social structure and values of a community. Te Hiku Media's work represents a challenge to the growing power of these tech giants, emphasizing the need for local control over cultural resources and the preservation of indigenous knowledge.

Ethical AI also involves addressing biases and ensuring that AI systems are fair and inclusive. AI systems trained on data from predominantly Western or industrialized sources often fail to account for the needs and perspectives of marginalized communities. By retaining control over the Māori language data, Te Hiku Media is ensuring that their language models are developed in a way that is inclusive and fair, giving the Māori people the tools to shape their own future.

This case study delves into the critical intersection between financial mathematics, ethical AI practices, and the societal and cultural impacts of data exploitation. Through the lens of Te Hiku Media's work, we explore how mathematical models can be applied to safeguard indigenous data, protect privacy, and ensure fairness in the rapidly evolving field of AI and machine learning.

he core issue at hand is the idea of 'data colonization'—where big tech companies extract and monetize data from indigenous communities without offering equitable returns. In this context, financial mathematics can be an important tool in modeling the risks and rewards of data ownership, while ensuring that the economic benefits of such data flow back to the community. Many tech companies profit from massive datasets, but the local communities—like the Māori people—often don't see the returns on these datasets. If Te Hiku Media were to sell Māori language data to multinational tech companies, there would be the risk of this rare and private linguistic data being lost, as it would become commercialized for profit rather than serving its original social and cultural purpose.

Putting economic considerations aside, another profound yet positive source of Te Hiku's confidence—as highlighted in our podcast —is what Keoni refers to as an "impressively low" error rate achieved through the organization's own efforts [1]. Machine learning operates through a programmed mechanism in which designers supply examples that show the desired output for a given input. Unlike typical NLP systems, which require extensive pre-training on large text corpora [2], Te Hiku is developing a new language tool based on semi-supervised learning [3], significantly reducing the need for labeled data. This approach enables Māori communities to revitalize their language independently, without relying on major technology companies.

When it comes to the assessment of the impacts, this act has and will exert a positive social influence on the completeness, continuity of Māori language and social fairness as well. Māori language suffered restriction imposed by primary education teachers in the 1970s, which has led to a

DOI: 10.54254/2754-1169/2025.BJ30766

significant loss of Māori speakers [3]. Now with the non-profit organization's rising attention to the language itself, this study is poised to spark attention, encourage study, foster development and drive the inheritance of this rare language among all age groups.

Moreover, from a national and social equity perspective, an independent non-profit language data initiative—rather than dependence on third-party technology interventions—can significantly improve educational equality for the Māori community. Language serves as the most fundamental and direct medium of communication, requiring continuous acquisition from early childhood [4]. When reliant on external data providers, local Māori speakers or those needing to learn the language could be subjected to first-degree price discrimination based on learners' socioeconomic background, leading to inherent educational inequity. This disproportionately affects innocent children, who bear no responsibility for their circumstances at birth [5]. In contrast, a locally managed non-profit can operate under principles of supervision and accountability that help ensure maximum fairness and protection.

The ongoing and substantial efforts by Te Hiku highlight the considerable relational, economic, and even political risks that may arise if language technology governance is mishandled. One crucial lesson in managing AI effectively is upholding data privacy and security. In recent years, numerous data monopoly incidents—such as Facebook's dominance in online advertising—have sparked public concern. Therefore, a key aspect of responsible AI development is ensuring transparent and ethical handling of personal data, including individual preferences and private information.

On a more optimistic note, this case represents a growing wave of courage and capability among local organizations—rather than tech monopolies—to lead meaningful change. The successes achieved by Te Hiku offer an inspiring precedent that may motivate broader resistance against persistent AI-related challenges, including algorithmic bias and exploitative economic uses of data [6].

5. Reflection

From a mathematical standpoint, statistical knowledge is essential for understanding this context. In particular, the Bayesian formula proves valuable for reasoning under conditional probabilities, while multivariate regression analysis helps manage and reduce potential errors, thereby mitigating biases that may arise during Te Hiku Media's own language data collection .

Although statistics aids in uncovering bias—and the ethical questions surrounding it—linear algebra plays an equally crucial role in speech-to-text technology. When systematically organizing, collecting, and analyzing language materials, applying core concepts such as differentiation and substitution through techniques like word embeddings allows linguists to map individual words or expressions into vector spaces. This reveals linguistic relationships and supports the construction of clearer and more coherent language structures. From a broader perspective, mathematical modeling serves as a bridge: it transforms complex real-world issues into well-defined, tractable mathematical problems.

The rapid development of generative artificial intelligence has continually transformed our understanding and daily experiences in recent years. However, beyond technological advances, the ethical questions it raises—which are deeply connected to philosophy—as well as the protective measures implemented by many institutions, warrant deeper reflection. The team's research analysis did not address ethical concerns related to collecting public data, even though a substantial number of the images were personal and sensitive [7]. Furthermore, since artificial intelligence interacts dynamically with human society, it may exert long-term influence on social structures and values. In

DOI: 10.54254/2754-1169/2025.BJ30766

this context, sociological knowledge becomes indispensable for addressing AI-related issues such as the Digital Divide and Social Justice, where vulnerable populations often face biased treatment.

In recent years, interdisciplinary research approaches have gained increasing attention. From my perspective, ethical decision-making modeling represents an ideal integration of mathematical modeling, statistics, and philosophy. Such models can be employed to evaluate the decisions made by language systems, ensuring they comply with ethical standards—including respect for individual rights and protection against privacy violations [8]. For example, mathematical models can help quantify the weighting of different ethical principles, allowing researchers to consider and evaluate multiple standards systematically.

Ultimately, the case of Te Hiku Media is a reminder that technology and culture need not be in opposition. With the right tools and approaches, communities can harness the power of technology to preserve and celebrate their cultures, protect their privacy, and ensure that their data remains a valuable resource for the community, rather than a commodity for profit [9]. As AI continues to shape the future, it is essential that we remain mindful of the ethical, social, and financial implications of technology, and work toward a more equitable, fair, and just world for all [10].

6. Conclusion

Te Hiku Media's work represents a profound integration of technology, ethics, and cultural preservation, demonstrating how Indigenous communities can reclaim authority over their linguistic and digital futures. By developing autonomous machine learning systems and adopting semi-supervised learning, the organization has reduced reliance on large labeled datasets—and thus on tech monopolies—enabling sustainable language revitalization on their own terms. Financially, the application of Bayesian modeling and multivariate regression allows for nuanced risk assessment and equitable decision-making regarding data usage, ensuring that economic benefits remain within the Māori community.

The case underscores the dangers of data colonization and the ethical necessity of protecting Indigenous knowledge from exploitative commercial practices. It also highlights how algorithmic sovereignty can prevent biased AI outcomes and promote educational equality, particularly for younger generations learning the language. Furthermore, Te Hiku's approach exemplifies the power of interdisciplinary collaboration—melding mathematics, sociology, ethics, and computer science—to address complex socio-technical challenges.

Looking forward, this initiative offers a scalable model for other marginalized communities seeking to harness AI ethically and autonomously. It reinforces the idea that technology should serve cultural and social goals—not undermine them. As AI continues to permeate society, ensuring that data governance remains transparent, inclusive, and community-led will be crucial in building a more equitable digital future. Te Hiku Media exemplifies how innovation can align with tradition, creating pathways for cultural survival in an increasingly data-driven world.

References

[1] Paris Marx, K.M. (2023) Tech won't save us: Big tech won't revitalize Indigenous Languages w/ Keoni Mahelona on Apple Podcasts, Apple Podcasts.
[2] Weinberg, J. (2022) Philosophers on GPT-3 (updated with replies by GPT-3) - daily nous, Daily Nous - news for & about the philosophy profession.
[3] Coffey, D. (2021) Māori are trying to save their language from Big Tech, WIRED UK.
[4] Floridi, L. and Chiriatti, M. (2020) 'GPT-3: Its nature, scope, limits, and consequences', Minds and Machines, 30(4), pp. 681–694.

DOI: 10.54254/2754-1169/2025.BJ30766

[5] Satz, D. (2012) 'Unequal chances: Race, class and schooling', Theory and Research in Education, 10(2), pp. 155–170. doi: 10.1177/1477878512446541.
[6] Wang, E. (2021) What does it really mean for an algorithm to be biased? The Gradient.
[7] Crawford, K. (2021) Atlas of AI: Power, politics, and the planetary costs of Artificial Intelligence. New Haven: Yale University Press.
[8] Fontes and Perrone (no date) Ethics of surveillance: Harnessing the use of live facial recognition ...
[9] Levy, F. (2018) 'Computers and Populism: Artificial Intelligence, jobs, and politics in the near term', Oxford Review of Economic Policy, 34(3), pp. 393–417. doi: 10.1093/oxrep/gry004.
[10] Liss, D. (2021) Today's real story: The facebook monopoly, TechCrunch.