

Accelerating the revitalisation of te reo Māori with AI

tehiku.nz // @tehiku
Peter Lucas Jones // @peterlucasjones
Keoni Mahelona // @mahelona



I ahu mai mātou i hea

- Te Reo Irirangi o Te Hiku o Te Ika (Te Hiku Media) Est. 1990
- Ngāti Kuri, Te Aupōuri, Ngāi Takoto, Te Rarawa, Ngāti Kahu.
- Te Tiriti o Waitangi, WAI 11, WAI 150, Wai 2224, Wai 262



Te Reo Māori Claim - Wai 11

Lodged by Huirangi Waikerepuru and Ngā Kaiwhakapūmau

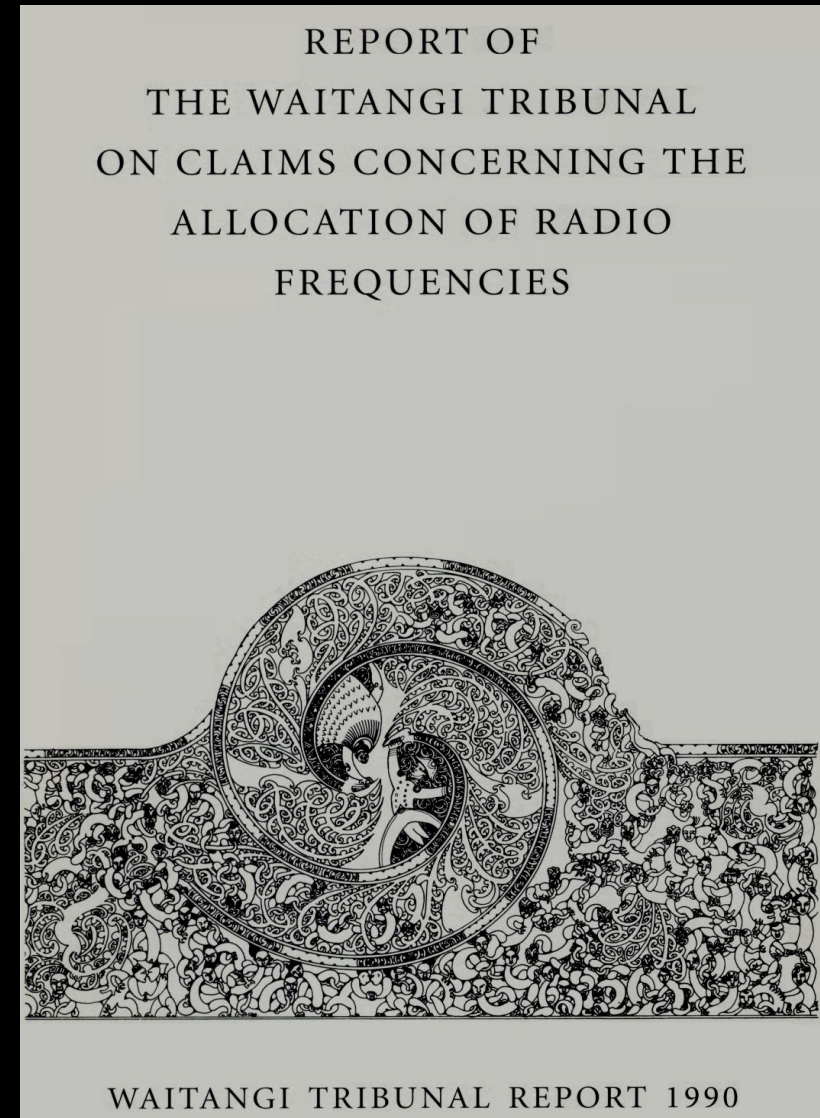
Tribunal Report published in 1986. It agreed te reo Māori was a taonga

The report made five recommendations. Tribunal recommendations are generally not binding, and the Crown implemented these recommendations in varying degrees.



In June 1990, claim Wai 150 was lodged by Sir Graham Latimer on behalf of the New Zealand Maori Council.

recognise and protect the Maori interest in radio frequencies



Rights to Radio Spectrum Wai 2224

- 3G, 4G, 5G
- Learn more at thtt.co.nz, Te Huarahi Tika Trust.
- Digital Switchover, 4G replacing terrestrial television, funded our content
- 4G treaty claim funded our speech recognition research

Te Reo Irirangi o Te Hiku o Te Ika

Vision

He reo tuku iho, he reo ora

Living language transmitted intergenerationally.

Mission

Whakatōkia, poipoia kia matomato te reo Māori o ngā
haukāinga o Te Hiku o Te Ika

Instill, nurture and proliferate the Māori Language unique to
haukāinga of Te Hiku o Te Ika

A color photograph of three elderly women sitting side-by-side on a light-colored couch. The woman on the left has short, curly white hair and is wearing a dark blue patterned top. The woman in the center has short, wavy white hair and is wearing a light pink patterned top. The woman on the right has short, wavy white hair and is wearing a colorful floral patterned top. A decorative fringe hanging from the ceiling is visible in the background.



30 years of corpus gathering trust and accountability



Tautinei

A collection of multimedia, written articles that promote and sustain public interest journalism

Latest

Resilience in Basketball amidst Pandemic
Iwi Government collaboration extinguishes Kaimaumau Fire
Remembering as an Act of Resistance

Live Broadcasts

Te Hiku Radio

Te reo irirangi o te hiku o te ika.



Recently Updated



Ngā Take

Ngā take o te Tai Tokerau.

5 hours ago in Te Reo



Kuaka Mārangaranga

Te Hiku talk about what matters in the kāinga, in the rohe and throughout the motu.

2 days ago in Te Hiku Radio



Pānui

Pānui for Te Hiku

4 days ago in Te Reo



Tōkihi Ki Tua 2021

Tōkihi ki Tua & Te Tai Tokerau Festival 2021



Ngā Take

Ngā take o te Tai Tokerau.

Latest

[Ngā Take o Te Taitokerau PM](#)

[Ngā Take o Te Tai Tokerau PM](#)

[Ngā Take o te Taitokerau PM](#)

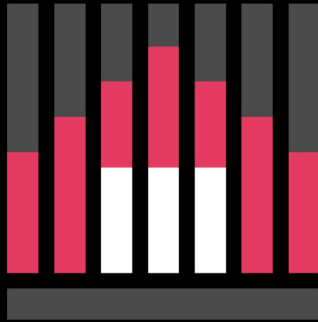
Live Broadcasts

[Te Hiku Radio](#)

Te reo irirangi o te hiku o
te ika.



https://whr.nz



Whare Kōrero

He mea tautoko nā



Carrier

10:34 PM



Nau mai ki te Whare Kōrero

Ko te mana raraunga Māori te tāhuhu o tēnei whare. Ka noho te mana o ngā taonga kei tēnei whare ki ngā kaitiaki, nō rātou anō ngā taonga. Tiakina ēnei taonga, kaua e takahi, kaua e pokanoa.

Welcome to the Whare Kōrero

This whare upholds the principles of Māori data sovereignty. All content remains under the guardianship of the original distributor. Respect this whare and the kōrero within it.

To learn more you can access the Kaitiakitanga License in the settings.

Tomo Mai

Carrier

10:35 PM



Kāinga



Te Hiku Media

Whakatōkia, poipoia kia matomato te reo Māori o ngā haukāinga o Te Hiku o Te Ika



Whakarongo
Te Hiku Media

Upcoming live streams:



Te Hiku Media

Tōkihi Ki Tua 2021

7:45AM • 11/05/2021



Remind
me

Recent uploads:



Kāinga



Latest



Rapu



< Te Reo o Te Kāinga Tuhinga Marau Hononga 9 Whakamārama

Āe.

Ko te rerekētanga o tēnā, tēnā kūmara, te **taputini**, arā, te, ka whakatongia. **He rerekē anō tōna whakatupuranga ki ngā kūmara e whakatupungia nei ināianei mō te hokohoko, he rerekē anō. Ko te ko ngā tikanga i whakatōngia i tēnā kūmara te taputini, nō muri kē i ngā mahinga, e, inā hauhakengia nē. Āe.**

Ko ngā mahi ka mahia hei a **akuaku** ngā **hiako** Āe, ki te, a, **hoi anō rā** i ngā wā o ngā mātua i ngā tūpuna, he mea, ko ngā **kota**. Ko ngā **kota**, āe, he aha? Te kūtai. Ko ngā **kota** kūtai, hei **akunga** ngā, ngā **taputini** nei, ā, he mā te āhua te tae o te **taputini** nei, ina **akungia** ana ngā **hiako** Nā, i te kāinga nei he kōhatu, he kōhatu tauhou tēnā ki te kāinga nei, kei roto i te, i ngā whārua o **Te Kao** nei, e **pēnei rā**, ehara tērā kōhatu nō te whenua e takoto nei te kōhatu nei, nē? Āe, he pēnei anō i te puke **keho** nei, e te āhua, engari tēnā kōhatu **Ko te mea kē**, he kānapanapa ana, nē? Āe, kānapanapa ana, tino rerekē tēnā te āhua o tēnā kōhatu, e, kia takoto i runga i te i

Taputini

Kupuake

Ko tētahi momo kūmara he taputini te ingoa i matomato te tupu i tēnei kāinga i Te Kao. E hara i te kūmara matatoru, engari he kūmara roroa. He mā tonu te āhua ā ko tōna tupu i rere kē atu i ērā kūmara arā e kore ōna rau, ōna tupu kūmara rānei e haere i runga i te whenua, heoi anō ka tupu tōpu noa iho. He kūmara reka, ā, he pai hoki mō te mahi hei kao.

na Toro Ihaka

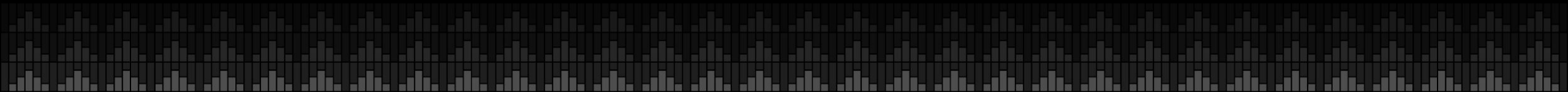


[https://](https://tku.nz/p.B2n)

tku.nz/p.B2n

Kōrero Māori

teaching computers indigenous languages



Kōrero Māori

Why

- Revitalisation, preservation, and promotion of te reo
- Create tech opportunities for indigenous people
- Transcribe native speaker radio archives

How

- Open source tools + Passionate Community ⇒
Automatic speech recognition

Kōrero Māori

hei whakaako i te rorohiko ki ngā reo taketake

KŌRERO

WHAKARONGO

Kōrero Māori on your mobile

Download the Kōrero Māori app on your mobile! You can read sentences even while you're offline.

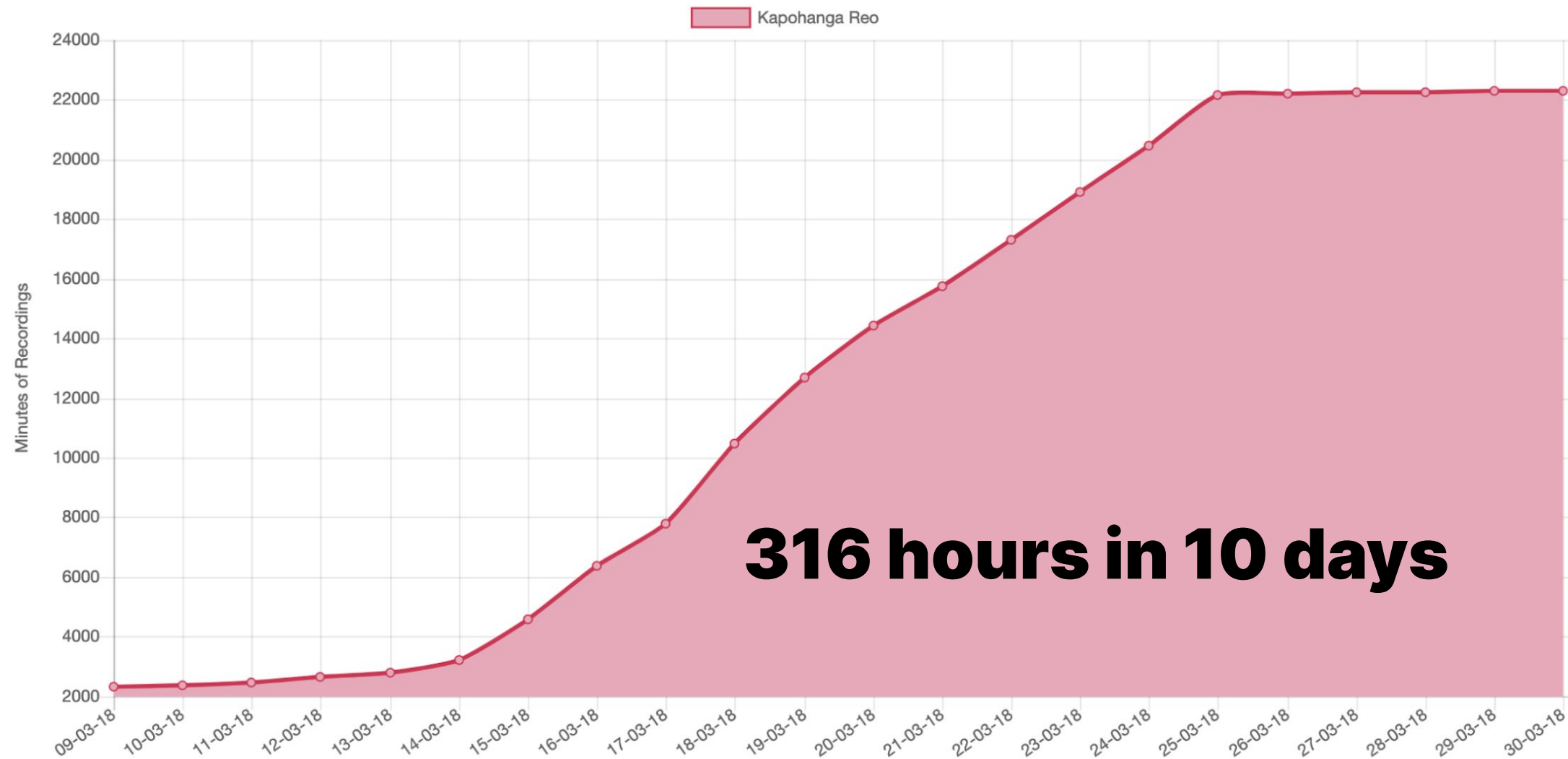


Rongorongo Competition

Congratulations to our winners!

The Rongorongo Competition has finished. Don't miss out on future competitions. Like our [Facebook Page](#) or [create an account](#) to stay in touch.





Stats on "corpus"

- Whare Kōrero
 - 3,500 hours of semi-labeled audio since November 2014
 - language, speakers, topics, description
- Archives
 - 50 VHS x 90 minutes
 - 600 DV tapes
 - 100s of cassette tapes
 - ~1600 recordings in total
- Kōrero Māori - 550 hours (utterances)

Using our corpus for revitalisation

Oro Tuku Iho

Intergenerational transmission of Sound

- E hoki whakamuri, kia anga whakamua
- The sound of our people, place and Identity
- Evolution vs Assimilation
- Decolonising the sound of our language.
Recognising the influence of English on pronunciation
- Speaking Native sound into our future.





Skip



Follow Up

Noise

Approve

Ka Tika

Kia Kaha

Delete

Incorrect Correct

Move the slider based on how good you think the speaker's pronunciation is.

Notes

Dataset None

H	e	o	i	,		p	i	r	i		a	n	a		a		T	e		A	r	a	w	a		k	i		t	e	
0	1	2	3	4	5	7	8	9	10	11	13	14	15	16	18	19	21	22	23	25	26	27	28	29	30	32	33	34	36	37	38

K	ā	w	a	n	a	t	a	n	g	a	,		t	a	n	a		o	a	t	i	t	i	a	n	g	a		m	ō	
40	41	42	43	44	45	46	47	48	49	50	51	52	54	55	56	57	58	60	61	62	63	64	65	66	67	68	69	70	72	73	74

t	a	n	a		r	o	h	e
76	77	78	79	80	82	83	84	85

Kaitiakitanga License

Kaitiakitanga License

- Uphold a set of principles that,
 - build Māori capability
 - ensure accountability and benefits back to the community from which the data/knowledge was derived
 - maintain sovereignty
- Prohibit use of data, software, tools etc. for applications that surveil, discriminate, persecute, etc.
- Affirmative action for data, software, etc.

Sovereignty for Indigenous Data & Tech

Digital Sovereignty

- Maintain sovereignty & guardianship over platforms and data
- Create opportunities for indigenous people
- Make indigenous language and culture more ubiquitous
- Promote language and culture from an indigenous world view

If we give corporates access to our data, they will profit from it and we will have the *privilege* of using their products and services.

Open doesn't always work for us

- Tragedy of the commons
- Privileged groups may benefit more from open/public goods
 - 18th century - take artifacts and put them in museums
 - 21st century - take data and put it in Common Voice
- National Parks
 - justify taking land from indigenous people for “public good”
 - “We’re better at managing natural resources”

Big Tech Colonialism/Imperialism

- Android or iOS
- Apple using "privacy" to maintain its walled garden
- In app & default browsers
- Sign in with _____
- xn--wharekrero-v3b.nz

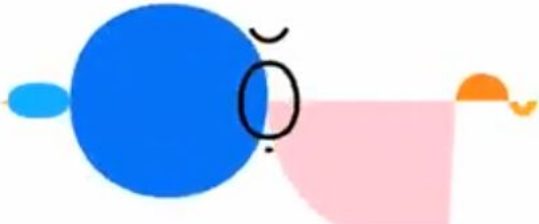


Facebook Viewpoints · Follow
November 6 at 4:02 AM ·

Join Facebook Viewpoints. Record phrases. Get rewarded.

Earn rewards


for recording phrases



0:01 / 0:03


Earn rewards for sharing your voice

INSTALL NOW



More people are learning Irish on Duolingo than there are native Irish speakers.

Welcome to the
Lionbridge Language Research
online registration!

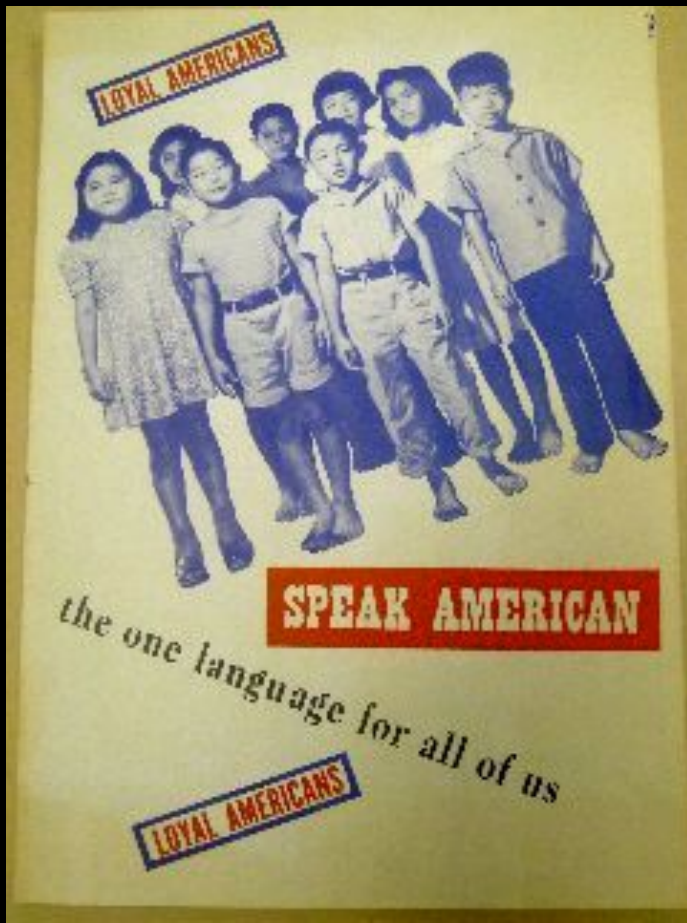


My name is Marc-Antoine, and I am part of the Lionbridge Artificial Intelligence Recruitment Team. I found your contact information from our internal database.

I am contacting you today because we have an ongoing Speech Data Collection project in New Zealand for adults and children who are native or fluent speakers of English that might interest you. The aim of the project will be to help improve a speech and ambient noise collection technology, as you might find

ancestryDNA
Where your story grows.





<https://www.duolingo.com> › course › hw › en › Learn-Hawaiian

The world's best way to learn Hawaiian - Duolingo

The world's most popular way to learn Hawaiian online Learn Hawaiian in just 5 minutes a day with our game-like lessons. Whether you're a beginner starting with the basics or looking to practice your reading, writing, and speaking, Duolingo is scientifically proven to work. Bite-sized Hawaiian lessons. Fun, effective, and 100% free.

Māori

English

kia kaha wāhine mā

let white women be strong

Open in Google Translate

Feedback

Māori

English

kiaora wahine mā

save white women

Open in Google Translate

Feedback

Ts & Cs

- "Free" content platforms e.g. <https://policies.google.com/terms?hl=en#toc-permission>
- Can't share data back to indigenous communities because of "copyright"

Right to Monetize

You grant to YouTube the right to monetize your Content on the Service (and such monetization may include displaying ads on or within Content or charging users a fee for access). This Agreement does not entitle you to any payments. Starting November 18, 2020, any payments you may be entitled to receive from YouTube

Scope

This license is:

- worldwide, which means it's valid anywhere in the world
- non-exclusive, which means you can license your content to others
- royalty-free, which means there are no monetary fees for this license

Rights

This license allows Google to:

- host, reproduce, distribute, communicate, and use your content — for example, to save your content on our systems and make it accessible from anywhere you go
- publish, publicly perform, or publicly display your content, if you've made it visible to others
- modify and create derivative works based on your content, such as reformatting or translating it

Language as a Service

Google ASR

Feature	Standard models (all models except enhanced video and phone call)		Enhanced models (video, phone call)	
	0-60 Minutes	Over 60 Mins up to 1 Million Mins	0-60 Minutes	Over 60 Mins up to 1 Million Mins
Speech Recognition (without Data Logging - default)	Free	\$0.006 / 15 seconds **	Free	\$0.009 / 15 seconds **
Speech Recognition (with Data Logging opt-in)	Free	\$0.004 / 15 seconds **	Free	\$0.006 / 15 seconds **

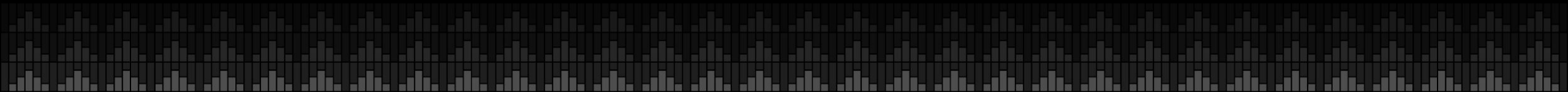
Who should profit from indigenous languages as a service?

Azure Services

Instance	Category	Features	Price
Standard - Web/Container 100 concurrent requests for Base model 20 concurrent requests for Custom model ¹	Speech to Text	Standard ²	\$1 per audio hour
		Custom	\$1.40 per audio hour Endpoint hosting: \$0.0538 per model per hour
		Conversation Transcription Multichannel Audio ^{PREVIEW}	\$2.10 per audio hour ³
	Text to Speech	Neural ⁴	Real-time synthesis: \$16 per 1M characters ⁴ Long audio creation: \$100 per 1M characters
		Custom Neural ^{4,5}	Training: \$52 per compute hour, up to \$4,992 per training Real-time synthesis: \$24 per 1M characters Endpoint hosting: \$4.04 per model per hour Long audio creation: \$100 per 1M characters
	Speech Translation	Standard	\$2.50 per audio hour
	Speaker Recognition	Speaker Verification	\$5 per 1,000 transactions
		Speaker Identification	\$10 per 1,000 transactions
		Voice Storage	\$0.20 per 1,000 voice profiles (10,000 free voice profiles per month)

A better future for AI

Rather than Big Tech gobbling up our data and selling it back to us, we should empower communities to lead their own platforms and solutions to help move their people forward.



Te Hiku Media's Achievements ML

- Super fast te reo Māori ASR
- Real-time pronunciation feedback on the edge
- Automatic POS
- Te reo Māori Speech Synthesis
- Kaitiakitanga License - getting global recognition in AI/ML





DONAVYN COFFEY

SECURITY 28.04.2021 06:00 AM

Māori are trying to save their language from Big Tech

Te Hiku Media gathered huge swathes of Māori language data. Corporates are now trying to get the rights to it



GETTY IMAGES / TIM GRAHAM / CONTRIBUTOR



ARTIFICIAL INTELLIGENCE

A new vision of artificial intelligence for the people

In a remote rural town in New Zealand, an Indigenous couple is challenging what AI could be and who it should serve.



Timnit Gebru ✓

@timnitGebru



This was a beautiful article to read after spending all weekend reading about the modern day eugenics movement underlying much of the whole AGI stuff.



Distributed AI Research Institute @DAIRInstitute · May 2

"Data is the last frontier of colonization," Mahelona says.

"In turning to AI to help revive te reo, the Māori language, Mahelona and Jones, who is Māori, wanted to do things differently."

technologyreview.com/2022/04/22/105...

[Show this thread](#)

12:59 PM · May 2, 2022 · Twitter Web App

Māori ASR Applications

[https://
reoora.com](https://reoora.com)

The banner features the 'REO ORA' logo at the top left, which includes a stylized Māori face. To the right is a navigation menu with links: Home, Latest Resources, About, Testimonials, and Contact. The central part of the banner is a dark blue area with white line-art icons and text. It includes 'SPEAK KŌRERO' with a microphone icon, 'LIVESTREAM CLASSES AKORANGA ARORANGI' with a group of people icon, 'TALK WHAKARONGO' with a speech bubble icon, 'WORDS KUPU' with a list icon, 'E-BOOKS E-PUKAPUKA' with a book icon, 'ONLINE' with a globe icon, 'TUIHONO' with a document icon, 'READ PĀNUI' with a book icon, and 'ALPHABET ARAPŪ' with a letter 'A' icon. The main text 'LEARN TE REO MĀORI' is written in large, colorful, block letters. Below this banner is a white box with the text 'Learn to Speak Te Reo Māori in 12 Weeks'.

REO ORA

Home Latest Resources About Testimonials Contact

SPEAK KŌRERO

LIVESTREAM CLASSES AKORANGA ARORANGI

TALK WHAKARONGO

WORDS KUPU

LEARN TE REO MĀORI

E-BOOKS E-PUKAPUKA

ONLINE

TUIHONO

READ PĀNUI

ALPHABET ARAPŪ

Learn to Speak Te Reo Māori in 12 Weeks

Kaituhi - ASR assisted transcription

He iti hau marangai ka tū te tīhokahoka

He iti hau marangai ka tū te tīhokahoka

Nāku te rourou nāu te rourou ka ora ai te iwi

Nāku te rourou nāu te rourou ka ora ai te iwi

Waiho ma te tangata e mihi

Mā tō rēo anō kia mārāma

E hoa mā, inā te ora o te tangata

Waiho ma te tangata e mihi

Mā tō rēo anō kia mārāma

Kaituhi

Automatic te reo Māori transcription

Kia ora e te whānau

Kia ora e te whānau

He iti hau marangai ka tū te tīhokahoka

Kia pai te rā

Mā tō rēo anō kia mārāma

He iti hau marangai ka tū te tīhokahoka

Go to Kaituhi

Kia ora e te whānau

Waiho ma te tangata e mihi

Mā tō rēo anō kia mārāma

Kia pai te rā

Nāku te rourou nāu te rourou ka ora ai te iwi

Nāku te rourou nāu te rourou ka ora ai te iwi

E hoa mā, inā te ora o te tangata

Nāku te rourou nāu te rourou ka ora ai te iwi

Mā tō rēo anō kia mārāma

E hoa mā, inā te ora o te tangata

Mā tō rēo anō kia mārāma

Kia pai te rā

He iti hau marangai ka tū te tīhokahoka

Mā tō rēo anō kia mārāma

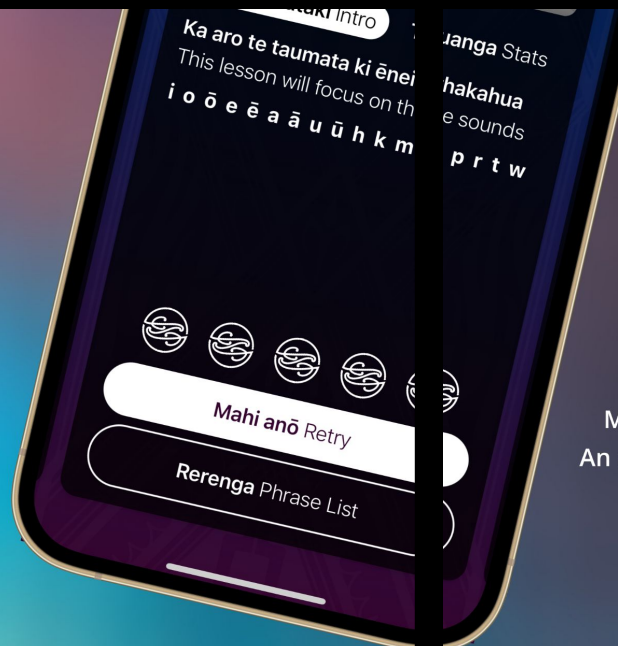
Waiho ma te tangata e mihi

Māori Pronunciation App

[https://
rongo.app](https://rongo.app)



Rongo



Improve your
Māori
language
pronunciation

Mā te rongo, ka mōhio
An aural first experience

Real-time
feedback

Practice in a private,
comfortable space
of your own



Natural Language Processing tools for te reo Māori.

tku.nz/p.G75

Links to articles and resources on data sovereignty

