CALEB MOSES

Indigenous Data Scientist

@ calebjdmoses@gmail.com ☎ +64 21 087 36640 📍 Auckland, New Zealand
🐦 @Caleb_Speak in calebjdmoses ○ mathematiguy

EXPERIENCE

Data Scientist
Dragonfly Data Science

📅 July 2018 – Ongoing   📍 Wellington, NZ

- Trained and deployed the first speech recognition algorithm for Māori language using deep learning in 2018
- Grew and trained a team of data scientists specialising in speech technology after securing $13 million of data science capability funding in 2019
- Wrote a custom toolkit for handling Māori language text for the purpose of NLP research in 2020
- Trained transformer-based language models and part of speech taggers for Māori language in 2021
- Developed a data processing and model pipeline system on top of Kubeflow to support a team of data scientists doing deep learning on prem

Statistical Analyst
Statistics New Zealand

📅 Jan - July 2018   📍 Wellington, NZ

- Worked on the System of Environmental and Economic Accounts (SEEA), a UN framework for environmental accounting
- As the resident R expert, trained new staff transitioning from SAS
- Implemented the first Natural Language Processing algorithm to be put into production for classifying building consents

Data Processing Analyst
Statistics New Zealand

📅 Aug 2016 - Jan 2018   📍 Wellington, NZ

- Subject matter expert on Confidentiality and Data Integration
- Pioneered the use of R to automate routine, manual confidentiality checks and improve scale and reliability
- Supported the Integrated Data Infrastructure: a national database of integrated administrative records from across government.

AFFILIATIONS

Member
Te Mana Raraunga: The Māori Data Sovereignty Network

📅 2018 - Ongoing   📍 NZ

Advocates for Māori rights and interests in data and algorithms in New Zealand

- Commented on behalf of the Te Mana Raraunga in Radio NZ article titled: Facial recognition regulations will be reviewed
- Co-authored press release on the use of facial recognition by NZ police

PROVERB

"Kāhore he aha e hangātia i ahu noa mai rānei, ka noho wehe i tēnei ao. Ahakoa, mata ngaro ka mōhiotia te mauri."

"Nothing in this world was ever created to live in isolation. Even a hidden face can be detected by its impact on its surroundings."

GOOD AT..

Ethics
Making connections with people to build technology that is an extension of the communities which it serves

Machine learning
Implementing, training, evaluating and deploying deep learning models for language applications

MLOps
Building scalable, reproducible pipelines to take data from raw, to model training, model evaluation and deployment on a cluster

Language
Speech recognition, speech synthesis, language modelling, text cleaning for multiple languages and data preparation for heterogeneous data

Corpus gathering
Building repositories to reproducibly collect, clean and store language data gathered using a variety of techniques including web scraping.

Research
Reading and synthesizing cutting edge research, re-orienting it to solve a new task, evaluating progress and reporting findings for a technical and non-technical audience

Public speaking
Explaining complex topics in simple terms, preparing attention-grabbing talks and even speaking with the media on occasion

Board member

Indigenous in AI

📅 Dec 2020 - Ongoing   📍 International

A Native, Aboriginal, and First Nations collective aiming to transform their home communities with advanced technology

- Co-presented on a NeurIPS 2021 panel titled: Indigenous Data Sovereignty
- Co-presented on a NeurIPS 2020 panel on Māori speech technology

Executive council member

NZ AI Forum

📅 Nov 2020 - Ongoing   📍 NZ

A not-for-profit organisation of AI industry leaders in NZ with the aim of growing the local AI ecosystem

- Aotearoa AI Summit 2021 invited panel member on the topic: Māori engagement within AI

Member

NZ AI Researchers Association

📅 Apr 2021 - Ongoing   📍 NZ

A not-for-profit organisation of AI researchers established to support the production and raise the standard of AI research in New Zealand

PUBLICATIONS

📘 Books

- Moses, C. (2020). Shouting zeros and ones: Digital technology, ethics and policy in new zealand. In A. Chen (Ed.), (Chap. The integrated data infrastructure. Te hanganga o ngā raraunga kōmitimiti me te whakaaetanga ā-iwi. pp. 84–101). Bridget Williams Books.

📄 Articles

- Lewis, J. E., Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., ... Moses, C. (2020). Indigenous protocol and artificial intelligence position paper.

👥 Conference Proceedings

- Moses, C. (2021). Language models in industry and around the world. In *Proc. of the forum "math for industry" 2021*.
- Moses, C., Mahelona, K., & Jones, P. (2021). Keynote: Papa reo - speech technology to support a bilingual aotearoa. In *Proc. of the language & society conference 2020*.
- Running Wolf, M., Arista, N., Running Wolf, C., Moses, C., & Davison, J. (2021). How to build-your-own practical a.i. tools for language maintenance. In *Proc. of the 7th international conference on language documentation & conservation (icldc 2021)*.
- Mahelona, K., Jones, P., & Moses, C. (2020). Progress in te reo māori speech recognition, speech synthesis, and real time pronunciation feedback. In *Proc. of the nz linguistics society conference (2019)*. Retrieved from https://papareo.nz/docs/PapaReo_NeurIPS2020_Poster.pdf
- Moses, C., Thompson, M., Mahelona, K., & Jones, P. (2019). Scoring pronunciation accuracy via close introspection of a speech recognition recurrent neural network. In *Proc. of neural information processing systems 32 (neurips 2020)*. Retrieved from https://papareo.nz/docs/PapaReo_NeurIPS2020_Poster.pdf
- Ramsden, A., & Moses, C. (2016). Redeveloping the confidentiality method for statistics new zealand business demography data. In *Proc. of the 2016 joint nzsa+orsnz conference*.

USER OF..

Docker   KubeFlow   AWS   emacs   git   Github

Python   PyTorch   TensorFlow   DVC   HuggingFace   pandas   scrapy   spacy   JupyterLab

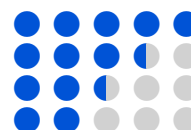R   brms   stan   dplyr   ggplot   Rmarkdown   data.table

LANGUAGES

English ●●●●●
Japanese ●●●◐○
Korean ●●◐○○
Te Reo Māori ●●○○○

EDUCATION

Postgrad diploma in Mathematics

University of Auckland

📅 2014-2015

Wrote dissertation titled: Potential Theory on Algebraic Curves in C^2

BSc. in Mathematics

University of Auckland

📅 2009-2013

Also studied Japanese Language, Philosophy, Physics and Economics

REFEREES

Dr. Edward Abraham
@ Dragonfly Data Science
✉ edward@dragonfly.co.nz

Dr. Finlay Thompson
@ Dragonfly Data Science
✉ finlay@dragonfly.co.nz

Assoc. Prof. Te Taka Keegan
@ University of Waikato
✉ tetaka@waikato.ac.nz

Mr. Andrew Sporle
@ University of Auckland
✉ a.sporle@auckland.ac.nz