

# Machel Reid

Website: <https://machelreid.github.io>

Email: [machelreid@weblab.t.u-tokyo.ac.jp](mailto:machelreid@weblab.t.u-tokyo.ac.jp) (UTokyo)

## EDUCATION

---

**University of Washington** *Currently taking gap year; 2022—Present; Seattle, WA*  
PhD Student (*Advisors: Luke Zettlemoyer & Noah A. Smith*) Expected Graduation: 2027/2028

## RESEARCH / WORK EXPERIENCE

---

**Google Research** London, UK  
*Research Scientist* Aug 2022 - Present

- Multilingual NLP, Low-resource NLP

**Matsuo Lab, The University of Tokyo** Tokyo, JP  
*Researcher* November 2020 - Aug 2022

- Working on multilingual pretraining methods, language modeling and text editing
- Teaching Assistant for undergraduate Introduction to Machine Learning (ItML) course.

*Research Intern* October 2019 - October 2020

- Working on representation learning, word embeddings for low resource languages, and definition modeling. Youngest author published in ICLR 2020, ACL 2020 workshops and at the main conference at EMNLP 2020, AAAI 2021.

**Language Technologies Institute, Carnegie Mellon University** Pittsburgh, PA  
*Visiting Student* May 2021 - August 2021

- Research internship advised by Professor Graham Neubig.
- Working on methods for modeling of editing processes

**Numada Lab, The University of Tokyo** Tokyo, JP  
*Research Intern* June 2019 - September 2019

- Worked on the development of a machine learning based system for accelerated disaster inquiry responses and for ease of access to information in natural disaster situations.

**Thomson Reuters Foundation** Remote / Tokyo, JP  
*Freelance* August – September 2018, May – June 2019

- Translated TRF's survey regarding social entrepreneurs and conducted research with companies across Japan to be incorporated into the global study. <http://poll2019.trust.org/>
- Conducted research about women-only carriages with Tokyoites in Japanese. Article featured on the World Economic Forum, Japan Times, Thomson Reuters. <https://www.weforum.org/agenda/authors/machel-reid>

## PUBLICATIONS

---

1. David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, **Machel Reid**, . . . ,Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. *The 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)*. July 2022. Association for Computational Linguistics.  
[Oral Presentation](#)
2. Itsuki Okimura, **Machel Reid**, Makoto Kawano and Yutaka Matsuo. On the Impact of Data Augmentation on Downstream Performance in Natural Language Processing. Workshop on Insights from Negative Results in NLP, ACL 2022. May 2022. Association for Computational Linguistics.  
[Oral Presentation](#) **Best Paper Award**
3. **Machel Reid**, Yutaro Yamada, Shixiang Shane Gu. Can Wikipedia Help Offline Reinforcement Learning? *January 2022. arXiv Preprint.*
4. **Machel Reid** and Graham Neubig. Learning to Model Editing Processes. *October 2021. Under Review*

5. **Machel Reid**, Junjie Hu, Graham Neubig, Yutaka Matsuo. AFROMT: Pretraining Strategies and Reproducible Benchmarks for Translation of 8 African Languages. *The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*. November 2021. Association for Computational Linguistics. [Best Paper Nomination](#), [Oral Presentation](#)
6. **Machel Reid** and Mikel Artetxe. PARADISE: Exploiting Parallel Data for Multilingual Sequence-to-Sequence Pretraining. *The 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)*. July 2022. Association for Computational Linguistics.
7. **Machel Reid** and Victor Zhong. LEWIS: Levenshtein Editing for Unsupervised Text Style Transfer. *Findings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP 2021 Findings)*. August 2021. Association for Computational Linguistics.
8. **Machel Reid**, Edison Marrese-Taylor and Yutaka Matsuo. Subformer: Exploring Weight Sharing for Parameter Efficiency in Generative Transformers. *Findings of The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021 Findings)*. November 2021. Association for Computational Linguistics.
9. Francis Zheng, **Machel Reid**, Edison Marrese-Taylor and Yutaka Matsuo. Low-Resource Machine Translation Using Cross-Lingual Language Model Pretraining. *AmericasNLP Workshop, NAACL 2021*. Association for Computational Linguistics.
10. Edison Marrese-Taylor, **Machel Reid** and Yutaka Matsuo. Variational Inference for Learning Representations of Natural Language Edits. *Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*. February 2021. Presented at the 5th Workshop on Representation Learning for NLP (non-archival), ACL 2020.
11. **Machel Reid**, Edison Marrese-Taylor and Yutaka Matsuo. VCDM: Leveraging Variational Bi-encoding and Deep Contextualized Word Representations for Improved Definition Modeling. *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*. November 2020. Association for Computational Linguistics.
12. **Machel Reid**, Edison Marrese-Taylor and Yutaka Matsuo. Combining Pretrained High Resource Embeddings and Subword Representations for Low-Resource Languages. *AfricaNLP Workshop, International Conference of Learning Representations (ICLR20) 2020*

## INVITED TALKS

---

### Incorporating Text Editing into Natural Language Processing

- Research Talk at the *CMU AI Seminar* (Host: Asher Trockman) 05/2022

### Data-centric techniques for improving low-resource multilingual pre-training and benchmarking

- Research Talk at the *Google Research* (Host: Clara Rivera) 04/2022

### On leveraging existing resources to develop reproducible benchmarks for African languages

- Research Talk at the *Masakhanne Machine Translation Mondays* (Host: Julia Kreutzer) 03/2022

### Edit Models & Learning and Understanding Implicit Structure in NLP

- Research Talk at the *The University of Tokyo*. (Host: Yoshihiro Kawahara) 12/2021

### PARADISE: Exploiting Parallel Data for Multilingual Sequence-to-Sequence Pretraining

- Research talk at *Facebook NLP Summit* (Host: Abhinav Arora) 10/2021

### Learning to Model Editing Processes

- Research talk at the *AIST* (Host: Edison Marrese-Taylor) 08/2021

## MENTORING

---

- **Itsuki Okimura** BS Student (2021-Present; now MS student) — studying impact of data augmentation techniques on finetuning pre-trained language models (for graduation thesis). Won the best paper at the Insights on Negative Results ACL 2022 workshop for this work.
- **Francis Zheng** MS Student (2021; now PhD student) — Using multilingual pre-training techniques for low-resource machine translation for Americas NLP Shared Task at NAACL 2021.

## AWARDS & HONORS

---

**NVIDIA Academic Hardware Grant** (PI: Edison Marrese-Taylor) Aug 2021  
Awarded single NVIDIA A100 for PCIe for academic research on Edit-Aware Language Models.

**Fugaku Project** (PI: Aleksandr Drozd)  
Awarded 1 million node hours on Fugaku<sup>1</sup> for “Training Novel Types of Large-Scale Language Models: Tuning”.

**Masayoshi Son Foundation Scholar** July 2020 - Present  
Supports youth with high aspirations and exceptional talents. Supports both full-tuition for higher education and research. Grants awarded Cross-lingual NLP research and for research on hierarchical language models (with Yutaro Yamada).

**Grand Prize Winner, Rakuten Technology Conference Hackathon** October 2018  
Developed a prototype matchmaking app between babysitters and working parents; made to combat problems that working women face in Japan today (career or children?). <https://blog.api.rakuten.net/rtc-hackathon-winners/>

**Matsuo Lab, SoTA Award** May 2021  
Awarded for exceptional contribution to the lab’s research.

## VOLUNTEER ACTIVITIES

---

- **Reviewing** AfricaNLP (2021,2022), ICLR (2022), ICML (2022), NeurIPS (2022), NAACL SRW (2022) ; Secondary: EMNLP (2020,2021), INLG (2020), EACL (2021), NAACL (2021)
- **Volunteer** ICML 2020, ACL 2020, ICLR 2020
- **Rakuten RapidAPI Champions Program**  
Worked in the capacity of a technology evangelist (API Champion) at Rakuten RapidAPI. Judged at Junction Tokyo 2019 Hackathon.

## SKILLS

---

**Software** Python, JavaScript, PyTorch, Fairseq, HF Transformers, Linux, Bash/Zsh, L<sup>A</sup>T<sub>E</sub>X  
**Languages** English (native), Japanese (advanced; almost native)

---

<sup>1</sup><https://www.fujitsu.com/global/about/innovation/fugaku/>