

# Andre Niyongabo Rubungo

770 Broadway  
New York City  
NY 10003, U.S.A

Email: rn3004 at princeton dot edu  
Homepage: <https://andrews2017.github.io/>  
Github: <https://github.com/Andrews2017>  
Twitter: [https://twitter.com/andre\\_niyongabo](https://twitter.com/andre_niyongabo)

## Education

Ph.D. in Computer Science, Princeton University Sep 2022 – Present  
Advisor: Adji Bousso Dieng

MS in Computer Science, University of Electronic Science and Technology of China (UESTC) Sep 2019 – Jul 2021  
Advisor: Hong Qu  
Thesis title: Text Classification for African Low-resource Languages: A case study of Kinyarwanda and Kirundi.

Bachelor in Computer Science, UESTC Sep 2015 – Jul 2019  
Thesis title: Deep Learning and Its Application in Text Classification  
Thesis advisor: Hong Qu

Chinese language, UESTC Sep 2014 – Jul 2015

## Employment

AI Research Intern May 2022 – Aug 2022  
Meta AI, New York, U.S.A  
Supervisor: Angela Fan

Research Assistant Oct 2021 – Aug 2022  
Polytechnic University of Catalonia, Barcelona, Spain  
Research group: MT@UPC

AI Research & Development Intern Jul 2021 – Sep 2021  
Huawei, Chengdu, China  
Supervisor: Wei Chen

AI Engineer Intern Jan 2021 – Apr 2021  
Whalecloud (Alibaba owned company), Guangzhou, China  
Supervisor: Chi Ma

## Publications

D. I. Adelani, J. O. Alabi, A. fan, et al. including **A. Niyongabo Rubungo**. *A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation*. Accepted at the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

Link: <https://openreview.net/pdf/ae270022e1f7523978bb7c46ec1706689f38ea38.pdf>

A. Nzeyimana and **A. Niyongabo Rubungo**. *KinyaBERT: A Morphology-aware Kinyarwanda Language Model*. Accepted at the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022.

Link: <https://aclanthology.org/2022.acl-long.367> **Best Linguistic Insight Paper Award**

I. Caswell, J. Kreutzer, et al. including **R. A. Niyongabo**. *Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets*. Transactions of the Association for Computational Linguistics (TACL), 2022.

Link: [https://doi.org/10.1162/tacl\\_a.00447](https://doi.org/10.1162/tacl_a.00447)

S. Gehrmann, T. Adewumi, et al. including **R. A. Niyongabo**. *The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics*. In the Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM), 2021.

Link: <https://aclanthology.org/2021.gem-1.10>

D. Moges, **R. A. Niyongabo**, and H. Qu. *Multi-Perspective Reasoning Transformers*. International Conference on Machine Learning and Computing (ICMLC), 2021.

Link: <https://dl.acm.org/doi/abs/10.1145/3457682.3457759>

D. I. Adelani, J. Abbott, et al. including **R. A. Niyongabo**. *MasakhaNER: Named Entity Recognition for African Languages*. Transactions of the Association for Computational Linguistics (TACL), 2021.

Link: [https://doi.org/10.1162/tacl\\_a.00416](https://doi.org/10.1162/tacl_a.00416)

**R. A. Niyongabo**, H. Qu, J. Kreutzer, and H. Li. *KINNEWS and KIRNEWS: Benchmarking Cross-Lingual Text Classification for Kinyarwanda and Kirundi*. The 28th International Conference on Computational Linguistics (COLING 2020).

W. Nekoto, V. Marivate et al. including **R. A. Niyongabo**. *Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages*. (EMNLP-Findings 2020. :Equal contribution), 2020.

Link: <https://www.aclweb.org/anthology/2020.findings-emnlp.195>

Y. Liu, W. Chen, **R. A. Niyongabo**, H. Qu, and K. Miao. *Weakly-supervised Character-level Convolutional Neural Networks for Text Classification*. the 14th International FLINS Conference (FLINS), 2020.

Link: [https://doi.org/10.1142/9789811223334\\_0084](https://doi.org/10.1142/9789811223334_0084)

F. A. Acheampong, H. Nunoo-Mensah, W. Chen, and **R. A. Niyongabo**. *Recognizing Emotions from Texts using a BERT-based Approach*. h International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2020.

Link: <https://ieeexplore.ieee.org/abstract/document/9317523>

## Conference Reviewing

International Conference on Learning Representations (ICLR) 2022

International Conference of the European Association for Machine Translation (EAMT) 2022

## Teaching

Teaching Assistant, University of Electronic Science and Technology of China (UESTC)  
Comprehensive Chinese Course Fall 2020

Teaching Assistant and co-organizer, University of Electronic Science and Technology of China (UESTC)  
The 1st and 2nd UESTC Deep Learning Workshops Fall 2018 and Spring 2019

Tutor, University of Electronic Science and Technology of China (UESTC)  
UESTC CS Web Development Tutorial Fall 2017 and Spring 2018

## Communication Skills

Human Languages: Kinyarwanda(Native), Chinese(Fluent), English(Fluent), Kirundi(Medium), French(Medium), Spanish(Beginner).

Machine Languages: Python (Pytorch, Fairseq, HuggingFace, Tensorflow, Numpy, Scipy, Sklearn, Pandas), SQL, JavaScript, CSS..

Version Control: Git.

Last updated: June 11, 2022