

Andre Niyongabo Rubungo

151 Taylor Ct
Princeton
NJ 08540, 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

Education

Ph.D. in Computer Science, Princeton University Advisor: Adji Bousso Dieng	Sep 2022 – Present
MS in Computer Science, Princeton University Advisor: Adji Bousso Dieng	Sep 2022 – March 2025
Pre-doctoral in Artificial Intelligence, Polytechnic University of Catalonia Advisors: Marta Ruiz Costa-jussa and Carlos Escolano Research focus: Machine Translation for Extremely Low-resource Languages	Oct 2021 – Aug 2022
MS in Computer Science, University of Electronic Science and Technology of China (UESTC) Advisor: Hong Qu Thesis title: Text Classification for African Low-resource Languages: A case study of Kinyarwanda and Kirundi.	Sep 2019 – Jul 2021
Bachelor's degree in Computer Science, UESTC (Li Ren class) Thesis title: Deep Learning and Its Application in Text Classification. Thesis advisor: Hong Qu	Sep 2015 – Jul 2019
Chinese Language Program, UESTC	Sep 2014 – Jul 2015

Employment

Research Scientist Intern (Agentic Coding Models Training) IBM Research, Yorktown Heights, NY, U.S.A Mentor & Manager: Juergen Bross	June 2026 – Sept 2026
Teaching Assistant Princeton University, New Jersey, U.S.A Courses: 1. Machine Learning and Data Science	Sep 2025 – Jan 2026
Research Intern (LLM-Granite 4.0 Post-training) IBM Research, Yorktown Heights, NY, U.S.A Mentor & Manager: Juergen Bross	May 2025 – Aug 2025
Teaching Assistant Princeton University, New Jersey, U.S.A Courses: 1. Mathematics for Numerical Computing and Machine Learning 2. Probability in Computer Science	Aug 2023 – May 2024

AI Research Scientist Intern (Hallucination Detection & Mitigation in MT Systems) May 2022 – Jul 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 Developer Intern Jan 2021 – Apr 2021
Whalecloud (Alibaba owned company), Guangzhou, China
Supervisor: Chi Ma

Publications

A. Niyongabo Rubungo, A. B. Dieng. *MOFSeq-2: Space-Group-Aware MOF Sequence Representations for Free Energy Prediction*. In preparation

A. Niyongabo Rubungo*, F. Fajardo-Rojas*, D. Gómez-Gualdrón, A. B. Dieng. *Highly Accurate and Fast Prediction of MOF Free Energy Via Machine Learning*. Journal of the American Chemical Society (JACS), 2025.

Link: <https://pubs.acs.org/doi/full/10.1021/jacs.5c13960>

A. Niyongabo Rubungo, C. Arnold, B.P. Rand, A.B. Dieng. *LLM-Prop: predicting the properties of crystalline materials using large language models*. npj Computational Materials, 2025.

Link: <https://www.nature.com/articles/s41524-025-01536-2> — **Best Poster Award at the Materials Research Society Conference, Fall 2023**

K. Li, **A. Niyongabo Rubungo**, X. Lei, D. Persaud, K. Choudhary, B. DeCost, A.B. Dieng, J. Hattrick-Simpers. *Probing Out-of-distribution Generalization in Machine Learning for Materials*. Nature Communication Materials, 2025.

Link: <https://www.nature.com/articles/s43246-024-00731-w>

A. Niyongabo Rubungo, K. Li, J. Hattrick-Simpers, A.B. Dieng. *LLM4Mat-Bench: Benchmarking Large Language Models for Materials Property Prediction*. Machine Learning: Science and Technology, 2024.

Link: <https://iopscience.iop.org/article/10.1088/2632-2153/add3bb>

A. Nzeyimana, **A. Niyongabo Rubungo**. *KinyaColBERT: A Lexically Grounded Retrieval Model for Low-Resource Retrieval-Augmented Generation*. ArXiv, 2025.

Link: <https://arxiv.org/abs/2507.03241> (Under Review)

O. Ogundepo, T.R. Gwadabe, C.E. Rivera, J.H. Clark, S. Ruder, et al. including **A. Niyongabo Rubungo**. *AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages*. Findings of the Association for Computational Linguistics: EMNLP 2023.

Link: <https://aclanthology.org/2023.findings-emnlp.997.pdf>

M.R.Costa-jussà, C. Basta, O Domingo, and **A. Niyongabo Rubungo**. *OccGen: Selection of Real-world Multilingual Parallel Data Balanced in Gender within Occupations*. In the proceedings of the 36th Conference on Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS), 2022.

Link: <https://openreview.net/pdf?id=tTPVefaATp6>

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*. In the Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.

Link: <https://aclanthology.org/2022.naacl-main.223/>

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

Link: <https://aclanthology.org/2022.acl-long.367> — *Awarded Best Linguistic Insight Paper, ACL 2022*

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

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 — *Best Paper Award at AfricaNLP Workshop, ACL 2021*

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).

Link: <https://aclanthology.org/2020.coling-main.480/>

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), 2020.

Link: <https://www.aclweb.org/anthology/2020.findings-emnlp.195> — *Wikimedia Foundation Research Award of the Year 2020*

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

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

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

Awards & Honors

Best Poster Award at the Materials Research Society Conference Fall	2023
---	------

Best Linguistic Insight Paper Award at ACL	2022
--	------

Best Paper Award at AfricaNLP, ACL Workshop	2021
Chinese Government Outstanding International Masters Student Award	2021
Wikimedia Foundation Research Award of the Year	2020
UESTC Outstanding Masters Student Award	2020
1st prize of Academic Achievement Award	2020
UESTC Outstanding Undergraduate Student Award	2018
1st prize of Academic Achievement and Excellent Performance Awards	2017
Winner of the UESTC Innovation and Creativity Coding Competition	2017
1st prize of Academic Achievement Award	2016
1st prize of Chinese Language Studies Award	2015

Professional Service and Leadership

Invited Talks

The North American Chapter of the Association for Computational Linguistics (NAACL), MIA Workshop	2022
The International Joint Conference on Artificial Intelligence (IJCAI), Diversity and Inclusion Event	2022

Journal Reviewing

Nature Computational Science
Transactions on Machine Learning Research (TMLR)

Conference Reviewing

International Conference on Learning Representations (ICLR)	2025, 2024, 2022
Conference on Language Modeling (COLM)	2024
International Conference on Machine Learning (ICML)	2025, 2024, 2023
International Conference of the European Association for Machine Translation (EAMT)	2023, 2022
International Conference on Neural Information Processing Systems (NeurIPS)	2022

Workshop Reviewing

Workshop on AI for Accelerated Materials Design (AI4Mat), NeurIPS	2025, 2024
Workshop on Natural Language Generation, Evaluation, and Metrics (GEM), ACL	2021

Workshop Organization

Bridging the Gap Between Practice and Theory in Deep Learning – ICLR https://sites.google.com/view/bgpt-iclr24#h.hlnai7sxxa1	2024
Natural Language Generation, Evaluation, and Metrics (GEM) – ACL https://gem-benchmark.com/team	2021

Coding Experience

Front-end developer, School of Computer Science Innovation Center, UESTC	2016 – 2019
--	-------------

Leadership experience

Academic Committee Member of International Student Union (ISU)	Sep 2020 – Jul 2021
Country Representative of Rwandan students at UESTC	Sep 2020 – Jul 2021
Academic Officer in charge of Undergraduate Students for ISU	Sep 2017 – Jul 2018
Representative of International Students in Chinese and Overseas Stud. Association (COSA)	Sep 2015 – Jul 2017

Professional Membership

Masakhane	Jan 2020 – Aug 2022
-----------	---------------------

Teaching

Teaching Assistant, Princeton University Probability in Computer Science	Spring 2024
Teaching Assistant, Princeton University Mathematics for Numerical Computing and Machine Learning	Fall 2023
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

*Other Skills**Communication Skills*

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

Technical Skills

Programming Languages: C/C++, Python, Linux Shell, Java.

Frameworks: Pytorch, Fairseq, HuggingFace, Numpy, Scipy, Sklearn, Pandas, BeautifulSoup, Selenium, Neo4J.

Web dev.: JavaScript, CSS, HTML, Bootstrap, PHP, SQL, XML/XSL.

Version Control: Git.

Last updated: September 9, 2026