

Ryandito Diandaru

ryanditodua@gmail.com | +62 812 3336 9604 | rayendito.github.io | github.com/rayendito

Education

Bandung Institute of Technology, Bandung, Indonesia

2019 – 2024

B. S. in Informatics/Computer Science advised by *Dr. Eng. Ayu Purwarianti, S.T, M.T.*

- Research: Machine Translation, Multilingual/Low-Resource NLP, Large Language Models (LLM)
- Thesis: Indonesian-Javanese XLM Machine Translation Aided with GPT-3.5 Data Augmentation
- Cum Laude; GPA: 3.68 / 4.00 ([Transcript](#))

Publications

* Indicates co-first authors

TDLE 2024 Could We Have Had Better Multilingual LLMs If English Was Not the Central Language?
(Oral Presentation) **Ryandito Diandaru**, Lucky Susanto, Zilu Tang, Ayu Purwarianti, and Derry Tanti Wijaya

SEALP 2023 Replicable Benchmarking of Neural Machine Translation (NMT) on Low-Resource Local Languages in Indonesia
Lucky Susanto*, **Ryandito Diandaru***, Adila Krisnadhi, Ayu Purwarianti, and Derry Tanti Wijaya

NeurIPS 2024 CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark
(In submission) David Romero, Chenyang Lyu, and **(74 others)**

EMNLP 2024 SEACrowd: A Multilingual Multimodal Data Hub and Benchmark Suite for Southeast Asian Languages
(In submission) Holy Lovenia, Rahmad Mahendra, and **(59 others)**

ACL 2023 NusaCrowd: Open Source Initiative for Indonesian NLP Resources
Samuel Cahyawijaya, Holy Lovenia, and **(46 others)**

Research/Work Experience

Boston University

Oct. 2022 -

Research Affiliate (NLP) under Derry Wijaya, Ph.D.

Jul. 2024

- Evaluated language models (XLM, ChatGPT, Llama2) on machine translation and evaluated synthetic parallel/bilingual data generation capabilities of ChatGPT.
- Analyzed machine translation scores (BLEU, COMET-22) against language similarity scores (URIEL database) to propose strategies for language data collection for future LLMs.
- Proposed methods to utilize resources beyond parallel sentences for machine translation.
- Curated monolingual and parallel resources of Indonesian low-resource languages.
- Wrote manuscripts and author rebuttals for international conference submissions.

Bandung Institute of Technology

Oct. 2022 -

Research/Programming Assistant (Freelance)

Jun. 2024

- Developed a module to split a corpus into sets that maximize dissimilarity and respective language data statistics for Aspect-Based Sentiment Analysis (ABSA) research.
- Developed a text classification pipeline using sentence vectorizers (fasttext, BERT) and classical ML algorithms to process agent reports. (NDA)
- Developed automation pipeline and dashboard for graduate students' thesis advising records.

AntaraETP Nov. 2023 -
Mar 2024
Machine Learning Engineer (Part-time)

- Developed a Retrieval Augmented Generation (RAG) system for fake news/hoax detection.
- Utilized regular expression patterns to process legal documents scanned with OCR.

Prosa.ai Jun. 2022 -
Aug. 2022
Software Engineer Intern

- Developed website (front-end) for various speech-processing services using NextJS Typescript, MaterialUI, i18next, and Lit web components.

Selected Projects

GPTnano — <https://github.com/rayendito/GPTnano>

A collection of neural network architectures, implemented *from scratch* in PyTorch. Implemented architectures:

- Decoder-only Transformers (GPTnano)
- RNN (Elman Simple Recurrent Network)

cmicrograd — <https://github.com/rayendito/cmicrograd>

An implementation of an autograd engine in C

Volunteering and Service

Student Volunteer IJCNLP-AAACL 2023, Bali, Indonesia	Nov 2023
Programming Mentor ARKAVIDIA (Informatics and IT Festival)	Sep. 2020 - Mar. 2021
Operational staff IEEE ITB Student Branch (Best staff November 2020)	Mar. 2020 - Mar. 2021
Data Science Bootcamp Organizer Masa Bina Cinta (MBC) HME ITB	Jun, 2021 - Aug. 2021

Awards and Certifications

Band 8.5 IELTS (International English Language Testing System) Jan. 2024

- Reading: **8.5**, Listening: **9.0**, Writing: **7.0**, Speaking: **9.0**. ([Certificate](#))

Finalist of the KORIKA Generative AI Innovation Competition Sep. 2023

- Developed *In Simple Terms*, an AI-generated children's encyclopedia using in-context learning heuristics to mitigate generative AI hallucination. ([Technical report](#))

Garuda ACE Research Scholarship Oct. 2022

- Advisor(s): *Derry Wijaya, Ph.D* (Boston University), *Dr. Eng. Ayu Purwarianti, S.T, M.T.* (Bandung Institute of Technology)
- Supported by the Indonesian Ministry of Education, Culture, Research, and Technology (MoECRT) ACE Open Research program.

Bronze Medal GEMASTIK XIV National Data Mining Competition Oct. 2021

- Entered by over 150 teams from all around Indonesia.
- Analyzed the sentiment of Indonesian social media posts about COVID-19 vaccines through classical machine learning models. ([Technical report](#))
- Utilized Indonesian pre-trained models (IndoBERT) for the final round (Question-Answering)

Skills and Languages

Technical Python (Pandas, Matplotlib, Sklearn, HuggingFace, PyTorch, OpenAI), JavaScript, TypeScript, Amazon Web Services (Familiar)

Languages Indonesian (native), English (fluent), Javanese (native), Spanish (limited)