

Blaise Cruz

Natural Language Processing Researcher

Algorithms and Research Lab (ARL)
Samsung Research Philippines

Email: me@blaisecruz.com

Website: blaisecruz.com

Github: [jcblaisecruz02](https://github.com/jcblaisecruz02)

Education

BS Computer Science

De La Salle University-Manila
Adviser: Dr. Charibeth Cheng
August 2015 to August 2019

Received the Most-outstanding Thesis Award for my thesis on Automatic Fake News Detection. Graduated with Honors. Academic Full Scholar.

Tools

Programming Languages

Python 3 (Preferred) and C/C++

Deep Learning

PyTorch 1.6+ (Preferred), TensorFlow 1.12, and JAX.

Toolkits

Fairseq, HuggingFace Transformers, OpenNMT-py, SpaCy, Moses, and NLTK.

ML and Data Science

Statsmodels, scikit-learn, Weights&Biases, Optuna, Jupyter, Numpy, Pandas, Scipy, Matplotlib, Seaborn

References

Mentors and supervisors who have written recommendations:

Dr. Charibeth Cheng

Associate Dean - CCS
De La Salle University-Manila

Dr. Arnulfo Azcarraga

Full Professor
De La Salle University-Manila

Dr. Daniel Stanley Tan

Postdoctoral Researcher
Academia Sinica, Taipei, Taiwan

Josen Daniel De Leon

Algorithms & Research Lab Head
Samsung Research Philippines

* More available upon request.

Experience

Samsung Research Philippines (SRPH)

Natural Language Processing Researcher — January 2021 to present

- Leads internal research on multilingual low-resource machine translation. Produced a system that ranked 6th in an international translation competition (WMT'21).
- Spearheaded SRPH's first industry collaboration project with an international AI company.
- Assisted in creating SRPH's specialized NLP research group, serving as founding head.
- Accelerated an internal NLU-related project that was suffering from lack of data to proof-of-concept stage within two months using my Low-Resource Language skillset.
- First-authored one international conference paper.

DLSU Center for Language Technologies (CELT)

Research Consultant for NLP and AI — January 2021 to present

Research Assistant — August 2019 to December 2020

- Led undergraduate research teams working on Dialogue Generation, Question Answering and Generation (QA/QG), NLI, and Fake News Detection. Three teams under my co-supervision have won the Most-Outstanding Thesis Award.
- Established the DLSU Machine Learning Group — the university's first Machine Learning-focused reading group composed of student leaders and mentors.
- Produced the Philippines' first large-scale pretrained Transformer models in Filipino and released benchmark datasets for various Filipino NLP tasks, the firsts of their kind.
- First-authored three international conference papers, one journal paper, and one preprint.

Other organizations I have been affiliated with for **research** and/or **internship** work:

- **DLSU Center for Complexity and Emerging Technologies** — Subject Matter Expert for Machine Learning and Data Science — August 2018 to August 2019
- **Senti AI** — AI Research Engineer Intern — April 2018 to August 2018
- **RCG Global Services PH** — Analytics & Big Data Intern — January 2018 to March 2018

Recent Publications

Here is a list of my most recent works. Names with an asterisk (*) denote equal contribution, with the ordering of the authors determined via a coin flip. See my website for the full list.

- Sutawika L* & **Cruz JCB***. (2021). Data Processing Matters: SRPH-Konvergen AI's Machine Translation System for WMT'21. To appear in *Proceedings of the Sixth International Conference on Machine Translation*.
- **Cruz JCB**, Resabal J, Lin J, Velasco D & Cheng C. (2021). Exploiting News Article Structure for Automatic Corpus Generation of Entailment Datasets. To appear in *Proceedings of the 18th Pacific Rim International Conference on Artificial Intelligence*.
- Lopez E*, Cruz D*, **Cruz JCB*** & Cheng C. (2021). Simplifying Paragraph-level Question Generation via Transformer Language Models. To appear in *Proceedings of the 18th Pacific Rim International Conference on Artificial Intelligence*.
- **Cruz JCB**, Tan JA & Cheng C. (2020). Localization of Fake News Detection via Multitask Transfer Learning. In *Proceedings of the 12th International Conference on Language Resources and Evaluation*.
- **Cruz JCB** & Cheng C. (2020). Evaluating Language Model Finetuning Techniques for Low-resource Languages. *Philippine Computing Journal* 15(1).