

LONG THANH MAI

Dublin, Ireland | (+353) 894671501 | mailong225@gmail.com

EDUCATION

University of Science, VNU-HCM

HCMC, Vietnam

Bachelor of Science in Computer Science

2013 – 2017

University College Dublin

Dublin, Ireland

Doctorate of Philosophy in Machine Learning

2020 – Present

Thesis: Towards human-like open-domain dialogue systems (submitted)

INDUSTRIAL EXPERIENCE

Huawei Ireland Research Centre

Dublin, Ireland

Research Intern

06/2022 – 09/2022

Improved factuality in abstractive text summarization

EM and AI

Remote

Natural Language Processing Engineer

10/2019 – 9/2021

Developed NLP solutions in a low-resource setting

- Collected, prepared, and annotated training data to support model development
- Trained, tested, and packaged models for text classification, speech recognition, speech synthesis, speaker diarization, and text normalization
- Reduced speaker diarization errors from 18% to 7% through end-to-end models and data augmentation
- Decreased speech recognition errors from 24% to 11% using self-supervised learning techniques

VNG

HCMC, Vietnam

Associate Software Engineer

4/2018 – 4/2019

Developed NLP tools to enhance a news recommendation system

- Prepared training data and built models for part-of-speech tagging, named-entity recognition, and news article clustering/tagging projects
- Improved news recommendation system click-through rate from 8% to 13%

TMA Solutions

HCMC, Vietnam

Software Developer Intern

07/2016 – 11/2016

Participated in research and development of mobile applications on Android for an e-commerce system

SKILLS & COMPETENCIES

- Programming languages: Python, Java
- Machine learning frameworks: Pytorch, HuggingFace, Scikit-learn
- Software development stacks: Unix, Github, AWS, GCP
- Web development stacks: Javascript, REST, Django, bs4 (scraping)
- Open source projects: <https://github.com/mailong25>

PUBLICATIONS

1. **Long Mai** and Bac Le: Joint sentence and aspect-level sentiment analysis of product comments on YouTube. **Annals of operations research** (2020)
2. **Long Mai** and Julie Berndsen: Unsupervised domain adaptation for speech recognition via error correction. **INTERSPEECH** (2022)
3. **Long Mai** and Julie Berndsen: I already said that! Degenerating redundant questions in open-domain dialogue systems. **ACL-SRW** (2023)
4. **Long Mai** and Julie Berndsen: Enhancing conversational quality in language learning chatbots: An evaluation of GPT4 for ASR error correction. **ICASSP** (2024)
5. Elena Shushkevich, **Long Mai**, Manuel V Loureiro, Steven Derby, and Tri Kurniawan Wijaya: SPICED: News similarity detection dataset with multiple topics and complexity levels. **LREC-COLING** (2024)
6. **Long Mai** and Julie Berndsen: Improving linguistic diversity of large language models with possibility exploration fine-tuning. Arxiv.org/abs/2412.03343 (2024)
7. **Long Mai** and Julie Berndsen: Real-time textless dialogue generation. In preparation (2025)

FUNDING AND AWARDS

- SFI research training fellowship
- LCT Erasmus mundus scholarship (nominated)

REFERENCES

Professor Bac Le, PhD

Head of department of Computer Science, University of Science, VNU-HCM
181, 227 Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam
Email: lhbac@fit.hcmus.edu.vn

Professor Julie Berndsen, PhD

Full Professor, UCD School of Computer Science
University College Dublin, Belfield, Dublin 4, Ireland
Email: julie.berndsen@ucd.ie