POTSAWEE MANAKUL

pm574@cam.ac.uk | github.com/potsawee | potsawee.github.io

EDUCATION

University of Cambridge Cambridge, UK
Ph.D. Candidate in Information Engineering (Computer Science) *Oct 2019 – 2023/24 (expected)*
- Work in the Machine Intelligence Lab under the supervision of Prof. Mark Gales
- Research Focus: Generative Language Models, Summarization, Text Evaluation, and ML for NLP
- Recipient of the Cambridge International & St John's College Scholarship (Full funding for PhD)

University of Cambridge Cambridge, UK
B.A. and M.Eng. in Information and Computer Engineering *Oct 2015 – Jun 2019*
- **top 1%** in the first-year undergraduate examination
- **top 3%** in the second-year undergraduate examination
- **top 4%** in the third-year undergraduate examination
- **Honour with Distinction** in the fourth-year (Masters Program)

WORK EXPERIENCE

Amazon Berlin, Germany
Applied Scientist Intern *Sep 2021 – Mar 2022*
- Project: Weakly Supervised Video Summarization for Target Video Classification Task
- Community Shopping Science Team with Hani Al-Shater and Nam Khanh Tran
- A paper accepted at the Amazon Computer Vision Conference (ACVC) 2022

Machine Intelligence Lab, Engineering Department Cambridge, UK
Undergraduate Research Intern *Jun 2018 – Aug 2018*
- Developed deep learning models for grammatical error detection
- Worked on data augmentation methods to improve grammatical error detection systems
- Co-authored a paper accepted at ICASSP 2019

MediaTek Cambourne, UK
Software Engineering Intern *Jun 2017 – Sep 2017*
- Developed GUI tools for a 5G simulator using C#/.NET framework and SQLite database
- Designed an algorithm to detect a critical path in hard real-time systems

Healthera Cambridge, UK
Software Engineering Intern *Jun 2016 – Aug 2016*
- Designed and worked on the front-end part of the company's website

RECENT ML/NLP PROJECTS

1. **SelfCheckGPT**: https://github.com/potsawee/selfcheckgpt
   - Developed methods to detect hallucinations in large language models such as GPT-3
   - A discussion with the Washington Post about LLM Hallucination & SelfCheckGPT
     → Article is available: https://wapo.st/3IMgiZw
   - The dataset has been downloaded on HuggingFace more than 2,000 times
   - This work inspires NVIDIA's NeMo-Guardrails
2. **MQAG**: https://github.com/potsawee/mqag0
   - Developed automatic multiple-choice question generation and question answering systems
   - Models are open-sourced on HuggingFace, and they are downloaded more than 1,000 times
3. **LongSum**: https://github.com/potsawee/longsum0
   - Developed state-of-the-art automatic summary generation systems for long articles and podcasts
   - Podcast summarization system won 1st place at Spotify Podcast Summarization Challenge 2020
   - The work was published at ACL 2021

Selected Research Publications

1. **P. Manakul**, Y. Fathullah, A. Liusie, V. Raina, V. Raina, and M. J. Gales, "CUED at ProbSum 2023: Hierarchical Ensemble of Summarization Models", in **BioNLP Workshop** at ACL 2023.
   [Best-performing system on the Problem List Summarization shared task]
2. **P. Manakul** and M. J. Gales, "Sparsity and Sentence Structure in Encoder-Decoder Attention of Summarization Systems", in **EMNLP** 2021.
3. **P. Manakul** and M. J. Gales, "Long-Span Summarization via Local Attention and Content Selection", in **ACL** 2021.
4. **P. Manakul** and M. J. Gales, "CUED_SPEECH at TREC 2020 Podcast Summarisation Track", in Text REtrieval Conference (**TREC**) 2020.
   [Won 1st place in the Spotify Podcast Summarisation Track, out of 8 teams & 22 systems]
5. **P. Manakul**, M. J. Gales, L. Wang, "Abstractive Spoken Document Summarization Using Hierarchical Model with Multi-Stage Attention Diversity Optimization", in **Interspeech** 2020.
   [Best Student Paper Finalist]
6. Y. Lu, M. J. Gales, K. M. Knill, **P. Manakul**, L. Wang, Y. Wang, "Impact of ASR Performance on Spoken Grammatical Error Detection", in **Interspeech** 2019.

Professional Services

Reviewer

- ICML2023, InterSpeech2023
- IEEE Transactions on Audio, Speech, and Language Processing 2022, 2023

Undergraduate Teaching (Supervision at Cambridge)

- Third-year Undergraduate: 3F7 Information Theory and Coding
- Second-year Undergraduate: 2P7 Vector Calculus, Linear Algebra, Probability

Lab Demonstration

- MPhil in Machine Learning and Machine Intelligence: MLMI13 Introduction to NLP
- First-year Undergraduate: Engineering Lego Lab

Honours & Awards

Machine Learning Competition

- 1st place in the Medical Note Problem List Summarization at the BioNLP Workshop at ACL 2023
- 1st place in the Spotify Podcast Summarization Challenge at TREC 2020 Podcasts Track

Scholarships and Grants

- Cambridge International & St John's College Scholarship (Full funding for PhD) — 2019
- United Steel Companies Scholarship (Academic Prize, St John's College) — 2019, 2018, 2017, 2016
- Wright Prize (Academic Prize, St John's College) — 2019, 2018, 2017, 2016
- UARP (Undergraduate Academic Research Project, St John's College) — 2018
- Winifred Georgina Holgate Pollard Memorial Prize (University Prize) — 2017
- Earle Prize (St John's College Year Group Prize) — 2017
- BP 1st Year Prize (sponsored by BP, based on the 1st year exam) — 2016
- Gaskell Prize (St John's College Year Group Prize) — 2016

Awards

- Silver Medal from 45th International Physics Olympiad (IPhO) in Astana, Kazakhstan — 2014
- Gold Medal from 15th Asian Physics Olympiad (APhO) at National University of Singapore — 2014
- Bronze Medal from 14th Asian Physics Olympiad (APhO) in Bogor, Indonesia — 2013

Technical Skills

Programming: Python, Matlab, C++
Machine Learning Tools: PyTorch, TensorFlow, AWS

Updated: June 2023