HUNG QUOC TO

tqh262@gmail.com — LinkedIn — Github — Google Scholar — Kaggle

RESEARCH INTERESTS

LLM-based Multi-Agent Systems, Large Language Models (LLMs), LLM Test-Time Compute Scaling

PUBLICATIONS

Functional Overlap Reranking for Neural Code Generation
Hung Q. To, Minh H. Nguyen, & Nghi D. Q. Bui.
Findings of the Association for Computational Linguistics **(ACL 2024)**[pdf] [GitHub]

Better Language Models of Code through Self-Improvement
Hung Q. To, Nghi D. Q. Bui, Jin Guo, & Tien Nguyen.
Findings of the Association for Computational Linguistics **(ACL 2023)**[pdf] [GitHub]

RESEARCH EXPERIENCES

FPT Software AI Center Dec 2021 - Present
AI Research Resident
Academic advisor: Dr. Bui Duy Quoc Nghi

Topic: Multi-Agent, Large Language Models (LLMs), AI4Code, CodeLLMs.

Automatic Codebase Migration System *(Mar 2024 - Nov 2024)*

- Modeled waterfall-software-development-process-inspired codebase migration system. Decomposing complex migration process into logical sub-steps with **LLM-based Multi-Agent system**.
- Demonstrated end-to-end code repositories migration from Python to Javascript.

Functional Overlap Reranking for Neural Code Generation *(May 2023 - Apr 2024)*
Github Paper

- Developed SRank, a **LLM inference-time compute scaling** method for code generation by leveraging **self-consistency** in code execution.
- Introduced new metrics **functional overlap** to measure the consensus in functionality across function clusters, which plays a key role in majority voting in self-consistency.
- Published in **ACL 2024 Findings**. SRank achieved **32.9% and 6.1% improvement** on average over greedy-decoding and existing SOTA reranking methods, respectively. SRank **ranked #7** globally on HumanEval, most popular benchmark for code generation, as of Jan 2024.

CodeCapybara - An Open-Source Instruction-Tuned LLM for Code Generation *(Mar 2023 - May 2023)*
Github

- Led open-source LLM development for **code generation**, implementing full-parameter and LoRA fine-tuning. Fine-tuned 53K+ instruction-output pairs, surpassing LLaMA and Alpaca on HumanEval and MBPP-S benchmarks.
- Repository has been achieved **170 stars** on Github.

Better Language Models of Code through Self-Improvement *(Mar 2022 - Apr 2023)*
GitHub Paper

- Developed a data augmentation technique leveraging **knowledge distillation** without requiring addtional annotated data, leading to consistent and significant improvements in performance over teacher (i.e baseline) models on code summarization and code generation.
- Published in **ACL 2023 Findings**; achieved SOTA results and **ranked #1** on Microsoft's CodeXGLUE for code summarization as of Apr 2022.

AWARDS AND HONORS

Kaggle Competitions:

- **Gold Prize - Top 1%** - SIIM-FISABIO-RSNA COVID-19 Detection Challenge ($100,000 prize) (May 2021 - Aug 2021)

Last updated: February 26, 2025

    - The competition focuses on detecting and classifying COVID-19 cases in chest X-ray images by identifying lung opacities
- **Silver Prize - Top 5%** - Human Protein Atlas - Single Cell Classification Challenge ($25,000 prize) (Jan 2021 - May 2021)
    - The competition focuses on classifying proteins in single-cell images to understand protein expression patterns at the cellular level
- **Bronze Prize - Top 7%** - BirdCLEF 2021 - Birdcall Identification Challenge ($5,000 prize) (Apr 2021 - Jun 2021)
    - The competition focuses on identifying bird species from audio recordings of their calls and songs

Mathematics Competitions:

- **Gold Prize** - April 30th Traditional Olympic Competition in Mathematics, Vietnam (2016)
- **Silver Prize** - Southern Summer Camp Competition in Mathematics, Vietnam (2015)

TECHNICAL SKILLS

- **Programming Languages:** Python (Proficient), C++, JavaScript
- **Technologies:** Shell Scripting, Docker, Git, Unix CLI, LaTeX
- **Deep Learning Framework:** PyTorch, Huggingface
- **Distributed Training & Inference:** Distributed Data Parallel, Fully Sharded Data Parallel, DeepSpeed, Accelerator
- **Parameter-Efficient Fine-Tuning (PEFT):** LoRA, Prefix Tuning, Prompt Tuning
- **Compression Techniques:** Knowledge Distillation
- **Agent & LLM Application:** Langchain, Langgraph, LLM tool-calling, Diverse agentic architectures
- **LLM Post-training:** Instruction fine-tuning, CoT fine-tuning, RLHF (PPO, DPO)
- **Test-time Compute Techniques:** Diverse CoT prompting techniques, Search and selection from LLM generations, Iterative self-improvement & reflection

EDUCATION

Foreign Trade University 2017 - 2022
Ho Chi Minh City, Vietnam
Bachelor of International Business Administration

Last updated: February 26, 2025