

Aidan Yang

aidan@cmu.edu <https://cs.cmu.edu/aidan>

EDUCATION

- **Carnegie Mellon University** Pittsburgh, PA
PhD in Computer Science. Advisors: Ruben Martins and Claire Le Goues 2021 - 2025 (expected)
 - Research: machine learning, large language models (LLMs), program analysis, formal verification
- **Queen's University** Kingston, ON
BEng, Computer Engineering and Mathematics Dual Degree 2016-2020
 - Thesis: Modeling the loss function of generative adversarial networks (GANs) with Rényi information measures

ACADEMIA

- **Carnegie Mellon University** Pittsburgh, PA
Undergraduate Researcher (REUSE). Advisors: Ruben Martins and Claire Le Goues Summer 2020
 - Built a program synthesis pipeline for refactoring data-science APIs (e.g., Tensorflow, Pytorch, Dplyr)
 - Generated satisfiability modulo theories (SMT) constraints using NLP and deep learning models
- **Queen's University** Kingston, ON
Undergraduate Researcher. Advisors: Ying Zou and Ahmed E. Hassan May 2018 - May 2020
 - Performed a novel empirical study on 69,851 releases and 67.7 million user reviews for 2,232 apps
 - Built NLP models to establish links between evolving software artifacts achieving 79.8% accuracy

INDUSTRY

- **Amazon** Portland, OR
Applied Scientist Intern Summer 2023
 - Lead a research project within the Amazon AWS Systems Manager team, resulting in a patent filing for the Amazon Automated Reasoning Group (ARG)
 - Finetuned a 16B parameter LLM for Neural Machine Translation (NMT) on the Rust programming language
 - Established formal equivalence verification using property based testing (PBT) and bounded model checking (BMC) on LLM output
- **Microsoft** Redmond, WA
Machine Learning Research Intern Summer 2022
 - Conducted machine learning (ML) and formal methods (FM) research as part of the MSR RiSE group (programming languages, software engineering and formal methods for high-performance computing)
 - Implemented the AlphaZero reinforcement learning model on a custom combinatorial problem space: synthesizing Nvidia collective communication algorithms for parallel training
 - Sped up the previous Monte Carlo Tree Search algorithm by 70% by achieving 80% more GPU utilization on an Azure DGX-1 cluster
- **AMD** Markham, ON
Software Development Engineer May 2020 - May 2021
 - Developed High Dynamic Range (HDR) and Freesync features for GPU drivers in C++
 - Completed a data pipeline linking monitor specifications to AMD driver UI, automating 5% of QA engineers's manual testing

SELECTED PUBLICATIONS

- **LLMAO: Large Language Models for Test-Free Fault Localization**
Aidan Z.H. Yang, Ruben Martins, Claire Le Goues, Vincent Hellendoorn
IEEE International Conference on Software Engineering (ICSE), to appear 2024
- **An Empirical Study on Release Notes Patterns of Popular Apps in the Google Play Store**
Aidan Z.H. Yang, Safwat Hassan, Ying Zou, Ahmed E. Hassan
Journal of Empirical Software Engineering (EMSE), 2021
- **SOAR: A Synthesis Approach for Data Science API Refactoring**
Ansong Ni, Daniel Ramos, Aidan Z.H. Yang, Ines Lynce, Vasco Manquinho, Ruben Martins, Claire Le Goues
IEEE International Conference on Software Engineering (ICSE), pp. 112-124, 2021
- **Predicting Co-Changes between Functionality Specifications and Source Code in Behavior Driven Development**
Aidan Z.H. Yang, Daniel Alencar da Costa, Ying Zou
IEEE/ACM 16th International Conference on Mining Software Repositories (MSR), pp. 534-544, 2019

TEACHING

- TA for grad level course “Large Language Models and Applications” at CMU, Fall 2023
- TA for undergraduate level course “Data Structures and Algorithms” at Queen’s University, Winter 2018

PRESENTATIONS

- SOAR: Synthesis for Open-Source API Refactoring. Presented at: Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH 2020) Student Research Competition. November 2020
- SOAR: A Synthesis Approach for Data Science API Refactoring. Presented at: Carnegie Mellon University - Institute for Software Research. August 2020
- Predicting Co-Changes between Functionality Specifications and Source Code in Behavior Driven Development. Presented at: International Conference on Mining Software Repositories (MSR). May 2019

AWARDS

- Second Place at Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH 2020) Student Research Competition
- SIGSOFT CAPS Student Travel Award for ICSE 2019