Huy (Ken) Tu

LinkedIn: huytu | **Website:** kentu.us | **Github:** HuyTu7 | **Cell:** 919-961-8256 | **Email:** huyqtu7@gmail.com

Machine Learning engineer/researcher passionate about AI Fairness, NLP/NLU, Personalization, and Software Analytics. Driven to solve ambiguous, complex, and meaningful problems.

PROFESSIONAL EXPERIENCES

Meta Platforms, Inc.
New York, NY

RESEARCH SCIENTIST — Feb 2022 - Present

- Improved algorithmic and systems fairness across the IG recommender system through pioneering state-of-the-art (SOTA) Responsible ML practices as part of the Instagram Well-being Fairness team.
- Sourcing Fairness (0 to 1 project): designed, built, piloted, and scaled the offline simulation framework to efficiently test and measure the influence of proposed interventions on recommender systems.
- Formulated novel system fairness assessment methodologies and integrated with existing ML infrastructure to enable fairer real world outcomes and perpetuate Responsible ML practices at scale.

MACHINE LEARNING ENGINEER INTERN — May 2021 - August 2021

- Knowledge-Graph (KG) based Generic Entity Recognition (GER) in Feeds and Stories Content Understanding team.
- Utilized **Pytorch** to leverage KG and expand the current mention detection (MD) pipeline to GER which boosts the performance on *low-resource domains by 33%* while *performing statistically better than the production model* on MD task.
- Incorporated the SOTA decoding method (from EMNLP'20) to *identify nested entities more accurately, up to 18%*.

Amazon.com Services, LLC
Pittsburgh, PA

APPLIED SCIENTIST INTERN — May 2020 - August 2020

- Multilingual Natural Language Understanding in Alexa.
- Utilized **Tensorflow** to explore cross-lingual transferring and expand the current monolingual pipeline to multilingual which boost the performance of low-resource languages and production model to *15% with just 50K instances*.
- The work was integrated into production & was documented as a research paper for The Web Conference's WMS.

Pinterest Inc.
San Francisco, CA

MACHINE LEARNING ENGINEER INTERN — May 2019 - August 2019

- Users' interest recommendation: boost users' engagements & serve as features candidate generation for downstream functions (ads, homefeed, etc) by building a RecSys prototype from *300k+ users' activities* (via **Presto** & **Hive**)
- Found biases within the existing models (PCA, SVD, NMF, & ALS) and designed a new model that is more *diverse (55%)* and *relevant (33%)* with temporal supervised learning (SVM via **Scikit-learn** & LSTM via **Keras**).

Computer Science Department @ NCSU
Raleigh, NC

RESEARCH ASSISTANT — August 2018 – December 2021

- AI4SE: Researched & built tools that are human-focused/explainable AI to better software development.
- SE4AI: Conducted qualitative and quantitative studies to understand how SE processes/philosophies can improve AI.
- Projects: **(1)** reducing efforts for obtaining quality data for software analytics, **(2)** NSF SI^2 applying empirical SE for computational science projects to improve software quality of non-traditional software development.

TEACHING ASSISTANT — August 2016 – May 2018

- Coordinated with the professor & other teaching assistants as a team to structure the course (SE, Parallel Architecture, AI, Data Structures), design tests, conduct review sessions, facilitate labs, and deliver the lesson effectively.

Mathematics & Computer Science Departments @ ASU
Boone, NC

UNDERGRADUATE RESEARCH ASSISTANT — August 2012 - August 2016

- Led and collaborated with professors on *6+ research projects* to prove mathematical theorems (e.g. graph theories and operation research), analyze statistical relationships, and prototype models (via **Python**, **Java**, **R**, and **MySQL**).

EDUCATION

NORTH CAROLINA STATE UNIVERSITY (NCSU) *Raleigh, NC*

Ph.D. in Computer Science Dec 2021

- Advisor: Dr. Tim Menzies (h-index=68) @ RAISE Lab (Real-world Artificial Intelligence for Software Engineering)

M.S. in Computer Science May 2019

APPALACHIAN STATE UNIVERSITY (ASU) *Boone, NC*

B.S. in Computational Mathematics, *magna cum laude* - GPA: 3.80 / 4.0 May 2016

PUBLICATIONS & RESEARCH PROJECTS

Fair-SSL (Semi-Supervised Learning): Building Fair ML Software with Less Data

- [Fairware 2022]. Fair-SSL applies 4 popular SSLers as pseudo-labelers to create fair models that require only 10% labeled data while achieving similar performance as 3 modern bias mitigation algorithms

DebtFree: Minimizing Labeling Cost in Self-Admitted Technical Debt Identification using SSL

- [EMSE journal 2022]. DebtFree starts with SE knowledge to pseudo-label the SATDs in the training data. Then, an incremental RF active learner identifies the remaining SATDs (reduce 99% of required data).

FRUGAL: Unlocking Semi-supervised Learning for Software Analytics

- [ASE 2021]. Incorporate SE knowledge to identity regions of interest (reduce 97.5% of required data).

Leveraging Multilingual Neural Language Models for On-Device NLU

- [The Web Conference's WMS 2021], as part of the Amazon 2020 internship.

Mining Scientific Workflow for Anomalous Data Transfers

- [MSR 2021], as part of NSF SI^2. An anomaly detector, X-FLASH, identifies faulty TCP signatures in Scientific Workflows (SW). X-FLASH outperformed the SOTA up to *40% relatively in recall within 30 iterations*.

Can you Explain that Text, Better? Comprehensible Text Analytics for SE Applications

- [Accepted for ICML's QAI 2021]. A tuned decision tree (d=4) on LDA topics that performs similarly to TFIDF+SVM.

Identifying Self-Admitted Technical Debts (SATDs) with Jitterbug: A Two-step Approach

- [TSE journal 2020]. Jitterbug separates SATDs as hard and easy TDs to find *close to 100% of easy TDs* while being able to find hard TDs more efficiently (with less human effort) than the prior state of the art methods.

Data Labelling with EMBLEM (and how that Impacts Defect Prediction)

- [TSE journal 2020], as part of NSF SI^2. A novel system with human + AI partnership (incremental SVM active learning) to label buggy commits *8 times faster* and help build defect predictors *78% more accurately*.

Is One Hyperparameter Optimizer Enough?

- [FSE's SWAN 2018] Empirical case study for hyperparameter tuning in software defect prediction.

AWARDS AND HONORS

ACM Grace Hopper and Richard Tapia, 2018-21 | Scholar

ACM Joint ESEC/FSE Keynote, 2018 | Keynote Co-author for Top-tier SE conference

Pi Mu Epsilon Mathematics Honor Society, 2013-Present | Academic Excellence, *top 5%* of the class

Student Employee of the Year, 2015-16

Graduate Merits Fellowship, 2015-16 | Notable Mathematics Graduate Student *($10,000+)*, ASU

Who's Who Among Students in American Universities, 2015-16 | National Recognition for Outstanding Leader

SERVICE

Research Program

IEEE EMSE & TSE Journal Reviewer 2019-20

East Coast Asian American Student Union (ECAASU)

Director Of Advocacy Summer 2016 – Summer 2018