

# Ken Tu

**Github:** [HuyTu7](#) | **LinkedIn:** [huytu](#) | **Website:** [kentu.us](#) | **Cell:** 919-961-8256 | **Email:** [hqtu@ncsu.edu](mailto:hqtu@ncsu.edu)

## SUMMARY

---

- Passionate about the synergy of human and AI to improve software development specifically and socio-technical ecosystems generally.
- Extensive individual & collaborative work in hyper-parameters tuning, NLP/NLU, and recommendation systems.

## EDUCATION

---

### **NORTH CAROLINA STATE UNIVERSITY (NCSU)**

*Raleigh, NC*

Ph.D. in Computer Science

Expected: Dec 2021

- Advisor: Dr. [Tim Menzies](#) (h-index=62) @ [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

## PROFESSIONAL EXPERIENCES

---

### **Computer Science Department @ NCSU**

*Raleigh, NC*

RESEARCH ASSISTANT

August 2017 - Present

- AI4SE: Research & build tools that are human-focused/explainable AI to better software development.
- SE4AI: Conduct qualitative and quantitative studies to understand how SE processes/philosophies can improve AI.
- Current Projects: (1) reducing efforts for obtaining quality data for software analytics, (2) [NSF SI<sup>2</sup>](#) applying empirical SE for computational science projects to improve software quality of non-traditional software development.

TEACHING ASSISTANT

August 2016 - Present

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

### **Facebook Inc.**

*Menlo Park, CA*

MACHINE LEARNING ENGINEER INTERN

May 2021 - August 2021

- Knowledge-Graph (KG) based Generic Entity Recognition (GER) for the Feeds & Stories Content Understanding team.
- Utilized **Pytorch** to leverage KG, expand the current mention detection (MD) pipeline to GER which boosts the performance of *low-resource domains recognition* by 33% while still performing statistically better than the production candidate model on MD task.
- Incorporated the state-of-the-art decoding method (from EMNLP) to *identify nested entities more accurately*, up to 18%. The project will be integrated into production.

### **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 published as a research paper for The Web Conference's MS.

### **Pinterest Inc.**

*San Francisco, CA*

MACHINE LEARNING ENGINEER INTERN

May 2019 - August 2019

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

## 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).

## University Housing @ ASU

Boone, NC

RESIDENT ASSISTANT (RA) & VP of RA COUNCIL

January 2013 - August 2015

- Organized educational and social cross-building wide programs that fostered community for 900+ students.
- Expressed the voice of resident assistants and students on campus through composing policy and legislation proposals from Resident Assistant Council to University Housing Leadership (13 proposals in 1 year).

## PUBLICATIONS & RESEARCH PROJECTS

---

Fair-SSL: Achieving Fairness using Semi-Supervised Learning - [Submitted to ICSE conference 2022]

DebtFree: A SSL Method for Recognizing Technical Debt - [Submitted to EMSE journal 2021]

The Changing Nature of Computational Science (CSc) - [Submitted to ICSME conference], as part of [NSF SI^2](#).

FRUGAL: Unlocking Semi-supervised Learning for Software Analytics

- [Accepted for ASE's 2021]. Incorporate SE knowledge to identity region of interests (reduce 97.5% of required data).

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.

Leveraging Multilingual Neural Language Models for On-Device NLU

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

Mining Scientific Workflow for Anomalous Data Transfers

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

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

- [Accepted for TSE journal 2020] Separated SATDs as hard and easy TDs to find close 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)

- [Accepted for 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 accurate.

Is One Hyperparameter Optimizer Enough?

- [Accepted for 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 (17% acceptance)

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

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 Student Leader

## SERVICE

---

### Research Program

IEEE EMSE & TSE Journal REVIEWER

2019-20

ICSE and FSE Conference STUDENT VOLUNTEER

2018-19

### Asian Students In Alliance @ NCSU

Raleigh, NC

COMMUNITY LIAISON

Summer 2017 - Summer 2019

### East Coast Asian American Student Union (ECAASU)

DIRECTOR OF ADVOCACY

Summer 2016 - Summer 2019