

# HARRY WANG

✉ hwang551@asu.edu · 🌀 DeviRule · 🏠 Homepage · 🏠 Tempe, AZ, USA

## 🎓 EDUCATION

---

<b>Arizona State University</b> , Tempe, Arizona	expected May 2027
<i>Phd student</i> in Computer Science GPA: 4.3	
<b>University of Southern California</b> , Los Angeles, United States	May 2022
<i>M.S.</i> in Computer Science GPA:3.6	
<b>University of California, Davis</b> , Davis, United States	June 2020
<i>B.S.</i> in Computer Science GPA:3.7	

## 👤 EXPERIENCE

---

<b>Arizona State University SEFCOM</b> Tempe, Arizona	Aug 2022 – Present
<i>PhD Student</i>	
<ul style="list-style-type: none"><li>• Leveraged machine learning and program analysis to identify vulnerability-inducing snippets in open-source C projects.</li><li>• Developed machine learning models using over 100,000 issues and pull requests from dozens of open-source projects to detect malicious intent, and deployed a website on AWS EC2 to visualize the data.</li><li>• Analyze those Open Source Projects and use tools like Angr, LLVM and AFL to perform static and dynamic analysis on these projects to find the vulnerability introducing commits that simplify the debugging process.</li></ul>	
<b>USC ITS Governance Risk and Compliance Team</b> Los Angeles, California	Oct. 2021 – May 2022
<i>Research scientist</i>	
<ul style="list-style-type: none"><li>• Developed and implemented key performance indicators to track and assess cyber security risk among all USC devices (e.g. servers and laptops) using Python.</li><li>• Created and presented four metrics to evaluate the efficiency of the USC IBM Bigfix patching process to USC IT Service management executives.</li><li>• Created a report using Pandas and Numpy to monitor the implementation of multi-factor authentication for over 100,000 USC email accounts.</li></ul>	
<b>SubSythn</b> Los Angeles, California	June. 2021 – Present
<i>Graduate Researcher</i>	
<ul style="list-style-type: none"><li>• Utilized statistical inference techniques to improve the accuracy of programming analysis tools by reducing false warning generation through symbolic execution.</li><li>• Enhanced symbolic execution performance by implementing under-constrained symbolic execution on a code corpus.</li><li>• Created control flow graphs of under-constrained symbolic execution using binary analysis tools such as KLEE.</li><li>• Integrated control flow graphs with a statistical inference system to enhance analysis capabilities.</li></ul>	
<b>FedNLP</b> Los Angeles, California	Sep. 2020 – June 2022
<i>Major Contributor</i>	
<ul style="list-style-type: none"><li>• Contributed to FedNLP, a <b>200 github star</b> library that supports federated learning in Natural Language Processing using Python and Pytorch</li><li>• Preprocessed data for 10 popular NLP datasets and provided general data loading API of these datasets for user to easily access them</li><li>• Implemented two data partition methods: Kmeans and LDA for datasets and use these two partition methods to construct Non-IID datasets for users to simulate real world scenarios so that users can use them to train federated machine learning models</li><li>• Integrated common distributed federated learning algorithms (e.g., FedAvg) for popular NLP models (e.g., LSTM, BERT) and provided distributed training API within the FedML framework</li></ul>	
<b>UC Davis DECAL Lab</b> Davis, California	Feb. 2019 – Jun. 2020
<i>Undergraduate Researcher</i>	
<ul style="list-style-type: none"><li>• Used LSTM Neural Network and Transformer Model to study the common coding expressions of human coding so that programming tools can apply this pattern to debug code or to generate robust and readable code</li><li>• Built a custom Lexer to transform the 20GB of Java code to tokens and output tokens to JSON files</li><li>• Extracted common coding expression such as "<math>i = i + 1</math>" from JSON files, fabricated less common expressions such as "<math>i = 1 + i</math>", and save them in the same context in JSON so that we can compare the results from deep learning models</li></ul>	

- Retrieved 20 GB tokens data from JSON files, applied word embedding to the tokens to get the word embedding vector and to train the 3 layer LSTM and Transformer Model
- Optimized models with Adam optimizer, Layer normalization and cross-validation which boost accuracy from 70% to 90%
- Visualized the results with ggplot2, concluded that human prefer certain expressions in coding and co-authored an academic paper listed below based on this finding

**LitePoint Corporation** San Jose, California

Jun. 2019 – Sep. 2019

*Summer Software Engineer Intern*

- Built an Extract Transfer Load(ETL) system that automatically transfers raw manufacturing data to Microsoft SQL Server database using Python
- Implemented an automation program which applied pattern matching algorithm to locate newly created or modified CSV files among 100GB CSV files within a minute
- Created a backend pipeline that automatically fetched 10GB of new CSV files from designated servers, performed data cleaning, and uploaded data into the database in less than 2 minutes using parallel load streams
- Optimized SQL query performance by 500% by creating indexes for frequently used tables, using optimizer hints, and refactoring obsolete SQL queries

## ♡ PUBLICATION

Casey Casalnuovo, Kevin Lee, **Hulin Wang**, Prem Devanbu, Emily Morgan "Do Programmers Prefer Predictable Expressions in Code?" accepted by Cognitive Science Dec 2020

Bill Yuchen Lin, Chaoyang He, Zihang Zeng, **Hulin Wang**, Yufen Huang, Mahdi Soltanolkotabi, Xiang Ren, Salman Avestimehr "FedNLP: A Research Platform for Federated Learning in Natural Language Processing" accepted by NAACL2022 findings April 2022

## 💡 HONORS

**UC Davis Dean's List** Fall 2019 - Spring 2020

**PLDI 2022 PLMW Scholarship** Summer 2022

**Arizona State University SCAI Doctoral Fellowship** Fall 2022

**Arizona State University Fulton Fellowship** Fall 2022

**PWN2OWN Toronto win \$20,000 prize on router category** Dec 2022

## ♡ PUBLICATION

Bill Yuchen Lin, Chaoyang He, Zihang Zeng, **Hulin Wang**, Yufen Huang, Mahdi Soltanolkotabi, Xiang Ren, Salman Avestimehr "FedNLP: A Research Platform for Federated Learning in Natural Language Processing" accepted by NAACL2022 findings April 2022

## ⚙️ SKILLS

- Programming Languages: C, Python, Ocaml, React, Typescript, SQL, Angr, LLVM