PINJIA HE

CNB H103.1,
Universitätstrasse 6,
8092 Zurich, Switzerland

(+41) 779984073
pinjia.he@inf.ethz.ch
https://pinjiahe.github.io

RESEARCH INTERESTS

I am passionate about **software engineering** and natural language processing. Specifically, I design practical and elegant reliability solutions for both traditional software and intelligent software. As part of my work, I have developed effective testing approaches for machine translation, realized an AI-powered software log management and analysis framework, and proposed QoS prediction techniques for service-oriented software.

EDUCATION

The Chinese University of Hong Kong 08/2013–03/2018
Ph.D. in Computer Science and Engineering
Supervisor: Prof. Michael R. Lyu, ACM & IEEE Fellow for software reliability

University of Illinois at Urbana-Champaign 07/2016–09/2016
Visiting Student
Supervisor: Prof. Tao Xie, IEEE Fellow for software testing and analytics

South China University of Technology 09/2010–07/2013
B.Eng. in Computer Science and Technology
Bilingual class (Rank: 2/30), where students were selected from the entire university via an exam on Math and English

WORKING EXPERIENCE

Postdoctoral Scholar, ETH Zurich 09/2018–present
Supervisor: Prof. Zhendong Su

Postdoctoral Scholar, The Chinese University of Hong Kong 03/2018–08/2018
Supervisor: Prof. Michael R. Lyu

Research Intern, Microsoft Research Asia 05/2014–08/2014
Software Analytics Group, Mentor: Prof. Hongyu Zhang

PUBLICATIONS

Highlights: ICSE(3), ESEC/FSE(1), ASE(2), TDSC(1), TPDS(1)
Google Scholar: citations: 1130, h-index: 15

(*Corresponding author)

Selected Publications

1. Pinjia He, Clara Meister, Zhendong Su. Testing Machine Translation via Referential Transparency. *International Conference on Software Engineering (ICSE)*, 2021. *Acceptance Rate: 22%*
2. Pinjia He, Clara Meister, Zhendong Su. Structure-Invariant Testing for Machine Translation. *International Conference on Software Engineering (ICSE)*, 2020. *Acceptance Rate: 21%*
3. Shashij Gupta, Pinjia He*, Clara Meister, Zhendong Su. Machine Translation Testing via Pathological Invariance. *ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2020. *Acceptance Rate: 28%*

4. Pinjia He, Jieming Zhu*, Shilin He, Jian Li, and Michael R. Lyu. Towards Automated Log Parsing for Large-Scale Log Data Analysis. *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 2018. *Impact Factor: 6.9*
5. Shilin He, Jieming Zhu, Pinjia He, and Michael R. Lyu. Experience Report: System Log Analysis for Anomaly Detection. *International Symposium on Software Reliability Engineering (ISSRE)*, 2016. *Acceptance Rate: 35%*
   Most Influential Paper (26 papers selected in 30 years)
   710+ stars and **260+** forks on GitHub

Other Publications

1. Jinyang Liu, Jieming Zhu, Shilin He, Pinjia He*, Zibin Zheng, Michael R. Lyu. Logzip: Extracting Hidden Structures via Iterative Clustering for Execution Log Compression. *International Conference on Automated Software Engineering (ASE)*, 2019. *Acceptance Rate: 20%*
2. Jieming Zhu, Shilin He, Jinyang Liu, Pinjia He, Qi Xie, Zibin Zheng, Michael R. Lyu. Tools and Benchmarks for Automated Log Parsing. *International Conference on Software Engineering (ICSE)*, SEIP Track, 2019. *Acceptance Rate: 25%*
3. Wenyu Wang, Wujie Zheng, Dian Liu, Changrong Zhang, Qinsong Zeng, Yuetang Deng, Wei Yang, Pinjia He, Tao Xie. Detecting Failures of Neural Machine Translation in the Absence of Reference Translations. *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Industry Track, 2019.
4. Pinjia He, Zhuangbin Chen, Shilin He, Michael R. Lyu. Characterizing the Natural Language Descriptions in Software Logging Statements. *International Conference on Automated Software Engineering (ASE)*, 2018. *Acceptance Rate: 17%*
5. Pinjia He, Jieming Zhu, Zibin Zheng, and Michael R. Lyu. Drain: An Online Log Parsing Approach with Fixed Depth Tree. *IEEE International Conference on Web Services (ICWS)*, 2017. *Acceptance Rate: 21%*
   Adopted by **IBM Could**: [blog post][IBM's code]
6. Jieming Zhu, Pinjia He, Zibin Zheng, and Michael R. Lyu. CARP: Context-Aware Reliability Prediction of Black-Box Web Services. *IEEE International Conference on Web Services (ICWS)*, 2017. *Acceptance Rate: 21%*
7. Jieming Zhu, Pinjia He*, Zibin Zheng*, and Michael R. Lyu. Online QoS Prediction for Runtime Service Adaptation via Adaptive Matrix Factorization. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2017. *Impact Factor: 2.6*
8. Jian Li, Pinjia He, Jieming Zhu, Michael R. Lyu. Software Defect Prediction via Convolutional Neural Network. *IEEE International Conference on Software Quality, Reliability and Security (QRS)*, 2017. *Acceptance Rate: 26%*
9. Pinjia He, Jieming Zhu, Shilin He, Jian Li, and Michael R. Lyu. An Evaluation Study on Log Parsing and Its Use in Log Mining. *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2016. *Acceptance Rate: 21%*
   520+ stars and **270+** forks on GitHub
10. Jieming Zhu, Pinjia He, Qiang Fu, Hongyu Zhang, Michael R. Lyu, and Dongmei Zhang. Learning to Log: Helping Developers Make Informed Logging Decisions. *International Conference on Software Engineering (ICSE)*, 2015. *Acceptance Rate: 19%*

11. Jieming Zhu, Pinjia He, Zibin Zheng, and Michael R. Lyu. A Privacy-Preserving QoS Prediction Framework for Web Service Recommendation. *IEEE International Conference on Web Services (ICWS)*, 2015. *Acceptance Rate: 20%*
12. Cuiyun Gao, Baoxiang Wang, Pinjia He, Jieming Zhu, Yangfan Zhou, and Michael R. Lyu. PAID: Prioritizing App Issues for Developers by Tracking User Reviews Over Versions. *IEEE International Symposium on Software Reliability Engineering (ISSRE)*, 2015. *Acceptance Rate: 19%*
13. Pinjia He, Jieming Zhu, Zibin Zheng, Jianlong Xu, and Michael R. Lyu. Location-Based Hierarchical Matrix Factorization for Web Service Recommendation. *IEEE International Conference on Web Services (ICWS)*, Industry Track, 2014. *Acceptance Rate: 30%*
14. Tong Zhao, Junjie Hu, Pinjia He, Hang Fan, Michael R. Lyu, Irwin King. Exploiting Homophily-based Implicit Social Network to Improve Recommendation Performance. *International Joint Conference on Neural Networks (IJCNN)*, 2014.
15. Jieming Zhu, Pinjia He, Zibin Zheng, and Michael R. Lyu. Towards Online, Accurate, and Scalable QoS Prediction for Runtime Service Adaptation. *International Conference on Distributed Computing Systems (ICDCS)*, 2014. *Acceptance Rate: 13%*

IMPACT

Machine Translation Testing Project: I am the lead of this project, which aims to automatically find errors in widely-used machine translation software such as Google Translate and Bing Microsoft Translator. Our project has successfully found **1000+** translation errors in Google Translate and Bing Microsoft Translator. The translation errors are diverse, including under-translation, over-translation, incorrect modification, word/phrase mistranslation, and unclear logic.

LogPAI Project: I am the lead of LogPAI, an open-source project towards AI-powered log analytics. We released a set of log datasets and open-source tools on GitHub, which together have been starred **2,000+ times**. The datasets have been **downloaded 20,000+ times** by more than **380 organizations** from both industry and academia. The tools and datasets have been used by many industrial practitioners from Huawei, Microsoft, IBM, VMWare, Samsung, Red Hat, Ericsson, MasterCard, Rapid7, Nvidia, China Mobile, China Unicom, and many others.

WS-DREAM Project: I was a core member of the WS-DREAM project, which is to disseminate open datasets and source code for Web service research. We have developed QoS prediction approaches for Web service recommendation. Our publications on WS-DREAM have been **cited 2000+ times** in total. Our WS-DREAM datasets have been downloaded by **370+ organizations** globally, and have been utilized in research by **200+ papers**.

SERVICES

I served as a PC Member in the following conferences:

- 2020: ECOOP Artifact, APSEC ERA
- 2019: ASE Demo

I served as a Reviewer in the following journals:

- 2020: TSE, TOSEM, EMSE, CSUR, JCST
- 2019: TSE, TOSEM, TKDE, SMC, IST
- 2018: TSE, TOSEM, TKDE, STVR, COSE

I served as an external/sub Reviewer in the following conferences:

- 2019: ICSE, FSE
- 2018: DSN, ICDCS, IJCAI
- 2017: ICDCS, ISSRE
- 2016: ICSE, DSN, WWW, KDD, ISSRE
- 2015: ICSE, WWW, NIPS, ISSRE, WSDM

I served as a main contributor to ACM SIGSOFT Empirical Standards.

TEACHING

Instructor
Research Topics in Software Engineering, Spring 2020 (ETH Zurich)

Teaching Assistant
Natural Language Processing, Fall 2020 (ETH Zurich)
Research Topics in Software Engineering, Fall 2020 (ETH Zurich)
Rigorous Software Engineering, Spring 2020 (ETH Zurich)
Software Engineering Seminar, Fall 2019 (ETH Zurich)
Research Topics in Software Engineering, Fall 2019 (ETH Zurich)
Rigorous Software Engineering, Spring 2019 (ETH Zurich)
Research Topics in Software Engineering, Fall 2018 (ETH Zurich)
Software Engineering, Spring 2016/2017 (CUHK)
Problem Solving By Programming, Fall 2013-2016 (CUHK)
Technology, Society and Engineering Practice, Spring 2015 (CUHK)
Computer and Society, Spring 2014 (CUHK)

MENTORING

Rohan Shah, intern at ETH Zurich, 11/2020-present (with Prof. Zhendong Su)

- Undergraduate student from IIT Bombay
- Topic: Testing Knowledge Graph

Shashij Gupta, intern at ETH Zurich, 05/2019-07/2019 (with Prof. Zhendong Su)

- Undergraduate student from IIT Bombay
- Topic: Testing Machine Translation
- Machine Translation Testing via Pathological Invariance. *ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2020.
- Second Prize in ACM Student Research Competition (ICSE), ACM, 2020.

Clara Meister, RA at ETH Zurich, 01/2019-02/2020 (with Prof. Zhendong Su)

- Master degree from Stanford University
- Topic: Testing Machine Translation
- Machine Translation Testing via Pathological Invariance. *ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2020.
- Structure-Invariant Testing for Machine Translation. *International Conference on Software Engineering (ICSE)*, 2020.

AWARDS AND HONORS

Most Influential Paper (26 papers selected in 30 years), ISSRE, 2019.

Excellent Teaching Assistant, The Chinese University of Hong Kong, 2016.

Visiting Research Awards, The Chinese University of Hong Kong, 2015-2016.

Travel Award for DSN, 2016.

Postgraduate Scholarship, The Chinese University of Hong Kong, 2013-2017.

Honorable Mention in the Mathematical Contest in Modeling (MCM), 2013.