Yicong Huang
Irvine, CA

☎ (510)-230-9932
✉ yicongh1@ics.uci.edu
in linkedin.com/in/yicong-huang/
⌂ github.com/Yicong-Huang

- Researcher in Distributed Systems & Database
- System builder for Machine Learning
- Coder with passion

EDUCATION

Ph.D., Computer Science, University of California, Irvine   **09/2019 — expected: 03/2025**
Lab: Information System Group   Advisor: Dr. Chen Li (chenli@ics.uci.edu)   GPA: 3.9/4.0
Research interest: Data Processing Systems, Database Management Systems (DBMS), Data-Intensive Scalable Computing (DISC) Systems

Bachelor of Science, Computer Science, University of California, Irvine   **09/2015 — 06/2019**

PUBLICATIONS (SELECTED)

Computer Science

[1] **Yicong Huang**, Zuozhi Wang, and Chen Li.
"Udon: Efficient Debugging of User-Defined Functions in Big Data Systems with Line-by-Line Control".
In: *Companion of the 2024 International Conference on Management of Data, SIGMOD/PODS 2024*. ACM, 2024.

[2] Zhihui Yang, Zuozhi Wang, **Yicong Huang**, Yao Lu, Chen Li, and X. Sean Wang.
"Optimizing Machine Learning Inference Queries with Correlative Proxy Models". In: *Proc. VLDB Endow.* 15.10 (2022), pp. 2032–2044.
URL: https://www.vldb.org/pvldb/vol15/p2032-yang.pdf.

[3] Zhihui Yang, **Yicong Huang**, Zuozhi Wang, Feng Gao, Yao Lu, Chen Li, and X. Sean Wang.
"Demonstration of Accelerating Machine Learning Inference Queries with Correlative Proxy Models".
In: *Proc. VLDB Endow.* 15.12 (2022), pp. 3734–3737. URL: https://www.vldb.org/pvldb/vol15/p3734-yang.pdf.

[4] Xiaozhen Liu, Zuozhi Wang, Shengquan Ni, Sadeem Alsudais, **Yicong Huang**, Avinash Kumar, and Chen Li.
"Demonstration of Collaborative and Interactive Workflow-Based Data Analytics in Texera".
In: *Proc. VLDB Endow.* 15.12 (2022), pp. 3738–3741.

[5] Avinash Kumar, Sadeem Alsudais, Shengquan Ni, Zuozhi Wang, **Yicong Huang**, and Chen Li.
"Reshape: Adaptive Result-aware Skew Handling for Exploratory Analysis on Big Data". In: *CoRR* abs/2208.13143 (2022).
DOI: 10.48550/arXiv.2208.13143. arXiv: 2208.13143. URL: https://doi.org/10.48550/arXiv.2208.13143.

Interdiscipline

[1] Judith Borghouts, **Yicong Huang**, Suellen Hopfer, Chen Li, and Gloria Mark.
"Wording Matters: the Effect of Linguistic Characteristics and Political Ideology on Resharing of COVID-19 Vaccine Tweets".
In: *Transactions on Computer-Human Interaction (TOCHI)* (2024).

[2] Joshua Rhee, **Yicong Huang**, Sadeem Alsudais, Shengquan Ni, Avinash Kumar, Chen Li, and David Timberlake.
"The marketing and perceptions of non-tobacco blunt wraps on Twitter". In: *Substance Use and Misuse* (2023).

[3] Judith Borghouts, **Yicong Huang**, Sydney Gibbs, Suellen Hopfer, Chen Li, and Gloria Mark.
"Understanding underlying moral values and language use of COVID-19 vaccine attitudes on twitter".
In: *PNAS nexus* 2.3 (2023), pgad013.

[4] Zimu Wang, **Yicong Huang**, Wanjun Lu, Jiaxin Liu, Xinying Li, Suhua Zhu, Hongbing Liu, and Yong Song.
"c-myc-mediated upregulation of NAT10 facilitates tumor development via cell cycle regulation in non-small cell lung cancer".
In: *Medical Oncology* 39.10 (2022), p. 140.

[5] Yawen Guo, Jun Zhu, **Yicong Huang**, Lu He, Changyang He, Chen Li, and Kai Zheng.
"Public Opinions toward COVID-19 Vaccine Mandates: A Machine Learning-based Analysis of U.S. Tweets".
In: *AMIA 2022, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 5-9, 2022*. AMIA, 2022.
URL: https://knowledge.amia.org/76677-amia-1.4637602/f006-1.4642154/f006-1.4642155/516-1.4642396/1066-1.4642393.

[6] Suellen Hopfer, Emilia J Fields, Yuwen Lu, Ganesh Ramakrishnan, Ted Grover, Quishi Bai, **Yicong Huang**, Chen Li, and Gloria Mark.
"The social amplification and attenuation of COVID-19 risk perception shaping mask wearing behavior: A longitudinal twitter
analysis". In: *PloS one* 16.9 (2021), e0257428.

[7] Lu He, Changyang He, Tera L. Reynolds, Qiushi Bai, **Yicong Huang**, Chen Li, Kai Zheng, and Yunan Chen.
"Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic".
In: *J. Am. Medical Informatics Assoc.* 28.7 (2021), pp. 1564–1573. DOI: 10.1093/jamia/ocab047.
URL: https://doi.org/10.1093/jamia/ocab047.

Blogs

[1] Aditya Verma, Dhruv Raipure, Jiaxi Chen, Sreetej Reddy, and **Yicong Huang**.
Enhancing the UDF Editor by Adding Language Server Support.
https://texera.github.io/blog/enhancing-the-udf-editor-by-adding-language-server-support/. 2023.

[2] **Yicong Huang**. *Using Texera to Perform Single-cell RNA Sequencing Analysis with R Language*.
https://texera.github.io/blog/using-texera-to-perform-single-cell-rna-sequencing-analysis-with-r-language/. 2023.

RESEARCH PROJECTS (SELECTED)

Texera ◯, Interactive Workflow-based Big Data Analytics System | Scala, Typescript, Python, Arrow   **09/2020 — present**
- Leading the effort on interactive debugging of Python UDF in the workflow-based Data-Intensive Scalable Computing (DISC).
- Designed and implemented the Python processing engine on top of Akka Actor system, targeting PySpark & PyFlink.
- Contributed in exploration of fault tolerence, version control on the workflow system.
- Intergrated progressive SQL adaptors on MySQL, PostgreSQL, AsterixDB.
- Maintaining a live service at Texera Prod

Cloudberry ◯, Big Data Visualization | Scala, Javascript, AWS   **06/2019 — 09/2020**
- Conducted tweet visualization with an interactive map, aggregating and displaying 4TB data stream.
- Integrated COVID-data with social media data on the interactive map.
- Built a fully scalable elastic service that can be load balanced on 20+ AWS machines, CoronavirusTwittermap.

ML-OPT, Machine Learning Pipeline Optimization | Python   **12/2019 — 06/2020**
- Explored using Machine Learning models to optimize Machine Learning Pipeline with a confidence guarantee.
- Conducted optimization on video recognition models (e.g., YOLOv3) by 25% with 98% accuracy guarantee, and NLP models (e.g., StanfordNLP) by 45% with 98% accuracy guarantee.

Wildfire, Wildfire Dectection & Visualization with Social Media | Python, Typescript, PostgreSQL   **06/2019 — 12/2019**
- Led team of 10 masters and undergraduates in detecting wildfires based on tweets and satellite data.
- Built data collection pipelines for real-time tweets, satellite data from NOAA, fire reports from USGS, etc.
- Integrated Machine Learning models such as AllenNLP, StanfordNLP, CNN, RNN, ANN for semantic analysis on text; ResNet50, VGG for images classifications.

Blockchain in Fin-tech & Blockchain and Smart Contracts | Python, C++   **07/2018 — 09/2018**
- Analyzed 200+ white papers for smart contracts that held on blockchains such as Ethereum, Bitcoin, etc.
- Built a prototype of smart contract on Ethereum.
- Arranged California-Shanghai Innovation Dialogue Conference 2018.

INDUSTRY EXPERIENCE

Research Intern | C++, Redis, RedisTimeSeries, Kafka, Flink   **06/2022 — 09/2022**
Visa Research, Visa Inc.   Palo Alto, CA
- Developed real-time window aggregation framework with out-of-order event support.
- Designed space-efficient, versatile list for in-memory raw-event storage.
- Proposed out-of-order handling algorithms, outperforming existing Flink designs.

Research Intern | C++, Kudu, MySQL   **06/2020 — 09/2020**
Infrastructure System Lab, ByteDance   Mountain View, CA
- Worked on an HTAP database to support instant query on the real-time data.
- Implemented TP (MySQL) metadata to AP (Kudu) schema conversion.
- Integrated lock-free data structure for heap implementation.

TEACHING EXPERIENCE

Research Mentor   **06/2019 — present**
University of California, Irvine   Irvine, CA
PhD students:
- **2022-2023:** Yunyan Ding (PhD), Xinyuan Lin (PhD), Raj Mohanty (PhD);
Master students:
- **2019-2020:** Yang Cao (MS);
- **2022-2023:** Aditya Verma (MS), Sreetej Reddy (MS), Dhruv Raipure (MS), Jiaxi Chen (MS);
Undergraduate students:
- **2019-2020:** Dayue Bai – UIUC (MS), Yinan Zhou – UCI (MS), Shiqi Wu – Berkeley (MS), Christine Xinrong Huang – CMU (MS), Tianran Liu – Univ of Washington (MS), Yutong Wang – UCD (PhD), Tingxuan Gu – CMU (MS), Yichi Zhang – NYU (MS), Xinyue Han – UCLA (MS), Qiaonan Huang (Hugo) – Brown (MS), Yuan Fu – CMU (SE-SV), Yuqi Huai – UCI (PhD), Quanzhen Du – UCSD (MS), Shiling (Scarlett) Zhang – Cornell (MS), Zeyad Kelani (MS);
- **2020-2021:** Chen He – CMU (MS), Bihao Xu – Uchicago (MS), Conghuai Tan (MS), Make Tao (MS), Mingshuo Liu (MS), Qifan Yu (MS);
- **2021-2022:** Zhen Guan – UCSD (MS), Jiashu Zhang – Hong Kong Polytechnic University (PhD), Yinan Zhou – UCI (PhD), Andrew Li (MS), Eric Peng (MS), Jiyang Wu (MS), Zeyu Li (MS);
- **2022-2023:** Chengxi Li (MS), Ethan Wong (MS), Tianyun Yuan (MS), Tony Liu (MS).

Teaching Assistant   03/2018 — 03/2022
University of California, Irvine   Irvine, CA

- ICS 51, Introduction to Computer Organization (W21);
- CS 122B, Projects in Databases and Web Applications (S18, W19, S19, S20, S21, F21);
- CS 141, Concepts of Programming Languages I (F18);
- CS 222/122P, Principles of Data Management (F19, W20, F20, W22).

Mentor   09/2018 — 12/2018
Dreams for Schools APPJAM+ | Yorba Linda High School   Yorba Linda, CA

Tutor   09/2016 — 03/2018
University of California, Irvine   Irvine, CA

- ICS 31, Introduction to Programming (F16);
- ICS 32, Programming with Software Libraries (W17);
- ICS 33, Intermediate Programming (S17);
- ICS 45J, Programming in Java as a Second Language (F17);
- ICS 46, Data Structure Implementation and Analysis (W18).

Volunteer Experience and Social Service

Student Volunteer
- California-Shanghai Innovation Dialogue Conference 2018   Irvine, CA
- VLDB 2019   Los Angeles, CA
- ICDE 2023   Santa Ana, CA
- SIGKDD 2023   Long Beach, CA

Guest Lecturer
- Review Sessions ICS 32 | CUCS 2018   Irvine, CA
- Review Sessions ICS 32, 33 | CUCS 2019 ***Best Lecturer Award**   Irvine, CA
- Application of Python UDF | DS4ALL 2023   Irvine, CA

Skills

Proficient	Python, Java/Scala, C/C++, JavaScript/TypeScript, Lisp, Prolog, MySQL, PostgreSQL
Intermediate	HTML, CSS, SQL, R, Redis, Golang
Programming Concepts	MapReduce, OOD, Functional, Logical Programming, Model-View-Control, Multithreading
Frameworks	Arrow, Hadoop, Spark, Flink, Kudu, Protobuf, Spring, Angular 2+, Django, Flask, Express, ReactJS, Elasticsearch
Services	AWS, GCP, GitHub, Docker