SHICHENG (GEORGE) LIU

CONTACT INFORMATION

1330 E. 53rd St., Chicago, IL, 60615 (+1) 773-236-6337
shicheng2000@uchicago.edu github.com/george1459

EDUCATION

(Honors) B.S. Computer Science & (Honors) B.S. Mathematics Jun. 2022 *(expected)*
Honors B.S. Computer Science with a specialization in Computer Systems
Minor in Physics
The University of Chicago, Chicago, IL
Cumulative GPA: 3.984/4.000

Quarter-long exchange at California Institute of Technology Autumn 2021
Exchange major: Computer Science
Exchange GPA: 4.1/4.3

PUBLICATIONS

Computer Science - Refereed Conference Paper

Automated Testing of Software that Uses Machine Learning APIs ICSE 2022
Chengcheng Wan, **Shicheng Liu**, Sophie Xie, Yifan Liu, Henry Hoffmann, Michael Maire, Shan Lu
44th International Conference on Software Engineering, 2022 (ICSE 2022)
(Acceptance Rate: 26.2%, 197 out of 751)

Are Machine Learning Cloud APIs Used Correctly? ICSE 2021
Chengcheng Wan, **Shicheng Liu**, Henry Hoffmann, Michael Maire, Shan Lu
43rd International Conference on Software Engineering, 2021 (ICSE 2021)
(Acceptance Rate: 22.4%, 138 out of 615)

ACADEMIC & RESEARCH EXPERIENCE

Undergraduate Researcher - ML API series study Jan. 2020 - Present
Prof. Shan Lu Research Group, Department of Computer Science, The University of Chicago

- Second-authored the ICSE 2021 paper (premier conference for software engineering) on Machine Learning (ML) API usage study. This is the first paper to systematically study and analyze applications that use cloud-based ML APIs. My specific contributions include:
  - Contributed to half of the selection of the non-trivial 360 benchmark applications from 1200+ open-sourced Github applications;
  - Identified and generalized 3 performance related anti-patterns through studying and profiling the use of ML APIs in these applications. Many of the mis-uses have been confirmed by developers;
  - Designed, implemented, and tested two static analyzers: Asynchronous API call checker and Constant-parameter API call checker. These checkers are capable of automatically identifying around 500 relevant Github Python applications as containing mis-uses;
  - Designed, implemented, and tested the speech recognition API wrapper;
- Second-authored the ICSE 2022 paper - our sequel work - on the first systematic testing framework for software systems that use cloud ML APIs. My specific contributions include:
  - Designed and implemented 10+ ML API reverse algorithms, crucial components of our tool algorithm. These include search-based inversions, synthesis-based inversions, and ML benchmark based inversions;

- ❍ Led a team development on a to-be-published VS Code plugin that visualizes the testing interface;
- ❍ Designed and implemented infrastructure code for large codebase and dependency support; evaluated tool effectiveness;
- ❍ Designed and conducted 10+ user study interviews and where developers utilize our tool to test software with ML APIs and provide feedback

Undergraduate Research Assistant - 3D Point Cloud Orientation Jun. 2021 - Aug. 2021
Prof. Rana Hanocka Research Group, Department of Computer Science, The University of Chicago

- Trained 3D point cloud segmentation neural networks in PointNet, a deep learning model for unstructured 3D point clouds;
- Incorporated network prediction result with dipole propagation and enhanced baseline approach to orient 3D point clouds

Astrophysics Research Intern - Classification of X-ray sources Jun. 2019 - Oct. 2019
Laboratory for Space Research, The University of Hong Kong, China
(UChicago Jeff Metcalf Intern)

- Integrated shell scripting pipelines to automatically gather and process Swift X-ray observation data from NASA database for Fermi 4FGL Gamma-ray catalog sources using NASA's HEASARC software;
- Developed Random Forest and Logistic Regression machine learning models based on known X-ray sources to characterize X-ray counterparts for unassociated 4FGL sources

Undergraduate Research Assistant - Objection detection on particles Jan. 2019 - Jun. 2019
Prof. William Irvine Lab, James Franck Institute, The University of Chicago

- Trained TensorFlow Object Detection API models, an image detection machine learning package, on Google Cloud. Developed and assessed models to detect experimental objects in particular fluids

TECHNICAL SKILLS

Python, C, C++, Rust, SQL, R, Julia, Haskell, Standard ML, Racket, Java/Type script, Bash/zsh
Linux, Git/SVN, PyTorch, Tensorflow, OpenGL

SELECTED HONORS & AWARDS

Grants & Scholarships (while @ UChicago, listed here)

College Summer Research Fellow ($5,000 grant) 2021
College Research Fellow ($4,500 grant) 2020-2021
Soong Ching Ling Foundation Scholarship ($12,500 scholarship) 2020
Jeff Metcalf Summer Research Fellowship ($6,000 grant) 2019

Academic Honors

Outstanding Undergraduate Researcher Award, Honorable Mention, CRA 2022

- CRA website notice
- UChicago CS News

Elected member of Phi Beta Kappa, the University of Chicago (the Beta chapter of Illinois) 2021
Enrico Fermi Scholar, The University of Chicago 2021
Robert Maynard Hutchins Scholar, The University of Chicago 2020
Dean's List, The University of Chicago 2018-2019, 2019-2020, 2020-2021

TEACHING EXPERIENCES

Teaching Assistant for CMSC 27200 Theory of Algorithms — Winter 2022

Grader for CMSC 22100 Programming Languages — Spring 2021

Teaching Assistant for CMSC 15100 Introduction to Computer Science I — Winter 2021

- TA and grader obligations are similar for these courses and both include: holding weekly office hours, answering piazza questions, and grading students' assignments and tests

SELECTED COMPUTER SCIENCE COURSE PROJECTS

Operating Systems	Implementation of the full Pintos OS kernel: (1) Thread scheduling (clock wait & priority scheduling with nested donation); (2) User program (stack setup, program wait/exit mechanism, system calls); (3) Virtual memory support (page table, frame table, eviction policy); (4) File system (index-structured inodes with doubly indirect pointers)
Introduction to Database Systems	Implementation of the full CrustyDB (an academic Rust-based relational DB): (1) Storage Manager (paging and file storage management); (2) Query Operators (aggregate, groupby, join fused in volcano-style interface); (3) Buffer Pool (buffer support for Storage Manager); (4) Performance Engineering; A Flask-based web DB server with a series of Entity-Relation and SQL designs
Introduction to Computer Systems	(1) Bit-level manipulation & arithmetic; (2) Debugging binary bombs by "deciphering" assembly lines; (3) Cache optimization with matrix transposition; (4) Simple shell (argument parsing, fork programs, output redirection); (5) Sample malloc library using segregated lists
Introduction to Machine Learning	(1) Regression using closed-form solution & gradient descent; (2) Logistic regression with softmax using gradient descent (applied to MNIST); (3) Modeling bias, variance, and noise to understand bias-variance tradeoff; (4) SVM using subgradient descent (applied to sentiment analysis); (5) Decision tree splitting by Gini index & random forest (applied to MNIST); (6) Neural network with forward and backward propagation
Natural Language Processing	(1) Implementation of various NLP models (implemented in *PyTorch*), specifically consisting of Logistic Regression Classifier, Deep Averaging Networks (DAN), and Sequence Neural Network Classifier (LSTM) with Attention and with BERT; (2) (Research Project) An investigation into characterizing relations between ideas, which applies topic modelling methods on computational social science. The final report is located at this GitHub repo
Computer Graphics	Implementation of a series of computer graphics projects, including from scratch: (1) Scene rendering using geometric transformations with NDC and line rasterization; (2) Triangle rasterization and the Gouraud and Phong shading algorithms; and implementations using OpenGL, the leading C++ API for graphics rendering: (3) Arcball rotation with Quaternions; (4) Phong shading with GLSL; (5) Implicit fairing for mesh smoothing; (6) keyframe interpolation with splines

Programming Languages	Implementation of program compilers, specifically consisting of the lexical tokenization, parsing into AST, desugaring, and evaluation phases of pedagogical languages, including both typed and untyped ones, and λ-Calculus
Parallel Computing	Implementation and systematic correctness and performance testing of: (1) A parallel version of the Floyd-Warshall algorithm; (2) Parallel work distribution using Lamport Queue with 1-to-1 worker-queue mapping; (3) Test-and-Set lock, Anderson's Array Lock, exponential Backoff Lock; (4) Load balancing using above locks with varying worker-queue mapping

PARTICIPATED SEMINARS

2022 UChicago Undergraduate Research Symposium *(scheduled)* May 2022

2021 UChicago Undergraduate Research Symposium May 2021

- Presented the ICSE 2021 paper in 2021 UChicago Undergraduate Research Symposium, the largest campus-wide interdisciplinary undergraduate research symposium. Poster available in the title link

8th Fermi Asian Network Workshop Summer 2019

- Participated in the 8th Fermi Asian Network Workshop in Zhuhai, China as summer intern at the University of Hong Kong. Program included (1) scientific talks related to Fermi-LAT gamma-ray observations and theoretical investigations and (2) workshops and training on Fermi-LAT data analysis

Undergraduate Reading Seminar in Particle Physics and Cosmology Winter 2019

LEADERSHIP & ACTIVITIES

Mentor for Student Summer CS Research Fellowship Program Jun. 2020 - Sept. 2020
Department of Computer Science, The University of Chicago

- Helped organize and coordinate weekly and end-of-program research presentations for a program consisting of 15+ international students conducting research at UChicago over summer

International Pre-Orientation Navigator (Orientation Leader) Sept. 2019
Orientation Leader Group, College Programming & Orientation Office, The University of Chicago

- Led a group of 10 first-year students through orientation programming, including campus and neighborhood tours, city-wide explorations, writing symposiums, and hosted an international student panel;
- Facilitated conversations and assisted incoming students in transitioning to college life at UChicago

RELEVANT COURSES

CMSC 16100 *(Honors)* Introduction to Computer Science I
CMSC 15200 Introduction to Computer Science II
CMSC 15400 Introduction to Computer System
CMSC 22100 Programming Languages
CMSC 23000 Operating Systems
CMSC 23010 Parallel Computing
CMSC 23500 Introduction to Database Systems
CMSC 25700 Natural Language Processing
CMSC 27100 Discrete Mathematics
CMSC 27200 Theory of Algorithms
CMSC 27410 *(Honors)* Combinatorics

CMSC 28000	Introduction to Formal Languages
CS 152	*(Caltech)* Introduction to Cryptography
CS/CNS 171	*(Caltech)* Computer Graphics Laboratory
TTIC 31020	*(PhD-level course)* Introduction to Machine Learning
MATH 16x00	*(Honors)* Calculus I-II-III
MATH 20x10	*(Accelerated)* Analysis in $\mathbb{R}^n$ I-II-III
MATH 20250	Abstract Linear Algebra
MATH 25x00	*(Honors)* Basic Algebra I-II-III
MATH 26500	Introduction to Riemannian Geometry
MATH 27000	Basic Complex Variables
Ma 109A	*(Caltech)* Introduction to Geometry and Topology
Ma 116A	*(Caltech)* Mathematical Logic and Axiomatic Set Theory
PHYS 14100	*(Honors)* Mechanics
PHYS 14200	*(Honors)* Electricity & Magnetism
PHYS 14300	*(Honors)* Waves, Optics, & Heat
PHYS 15400	Modern Physics
PHYS 18500	Intermediate Mechanics
PHYS 23x00	Quantum Mechanics I-II

ACADEMIC REFERENCES

Prof. Shan Lu shanlu@uchicago.edu
Professor, Department of Computer Science
The University of Chicago, Chicago, U.S.A.

Prof. Pablo Saz Parkinson pablosp@hku.hk
Research Assistant Professor, Department of Physics and Laboratory for Space Research
The University of Hong Kong, Hong Kong, China

This CV is last updated on Feb. 19th, 2021