# Yuen Chen

yuenc2@illinois.edu | chenyuen0103.github.io | chenyuen

## EDUCATION

**University of Illinois at Urbana-Champaign** — Expected May 2028
Ph.D. Student in Computer Science, advised by Prof. Hari Sundaram
Research Topics: Causality, Distributed Optimization, Domain Generalization, Transformer Architecture

**University of California Berkeley** — Graduated December 2022
B.A. Applied Mathematics & Statistics — GPA: 3.93/4.00

**University of Copenhagen** — February 2022 - June 2022
Exchange Student in Computer Science and Mathematical Science — GPA: 10.5/12

**Irvine Valley College** — Graduated May 2020
Associate in Science in Mathematics — GPA: 3.94/4.0

## SKILLS

**[Theoretical Domains]** Causal Inference, Machine Learning, Natural Language Processing
**[Programming Languages]** Python, R/R Studio, MATLAB
**[Machine Learning Frameworks and Libraries]** PyTorch, TensorFlow, NumPy, Pandas, Pyro

## RESEARCH EXPERIENCE

**Graduate Researcher | University of Illinois at Urbana-Champaign** — August 2023 - Present
*Advisor: Prof. Hari Sundaram*

- Conducting research on domain generalization to improve machine learning models' performance under distribution shifts
- Designing a dynamic distributed SGD scheme that optimizes task allocation among worker nodes based on computational speed
- Investigating the representation power and in-context learning ability of Transformers

**Research Intern | Empirical Inference, Max Planck Institute for Intelligent Systems** — February 2023 - July 2023
*Supervisors: Prof. Bernhard Schölkopf & Zhijing Jin (Ph.D. at Max Planck Institute & ETH)*

- Developed the "CLadder" benchmark to assess formal causal reasoning ability in Large Language Models (LLMs)
- Proposed a prompting-strategy to elicit multi-step formal causal reasoning in LLMs.
- Conducted experiments analyzing the impact of semantic representations on LLM performance

**Undergraduate Researcher | Language, Reasoning and Education Lab, ETH Zürich** — August 2023 - December 2022
*Supervisors: Prof. Mrinmaya Sachan & Zhijing Jin (Ph.D. at Max Planck Institute & ETH)*

- Conducted research on experimental design to optimize experimental tables with orthogonal arrays

**Undergraduate Researcher | University of Copenhagen** — February 2022 - June 2022
*Supervisor: Prof. Yevgeny Seldin*

- Conducted research to find an optimal algorithm for multi-armed bandits problems with time-based switching costs
- Reviewed literature in online learning, gaining insights into current methodologies and trends

## PUBLICATIONS

CLadder: Assessing Causal Reasoning in Language Models. *(NeurIPS 2023)*
Zhijing Jin*, **Yuen Chen***, Felix Leeb*, Luigi Gresele*, Ojasv Kamal, Zhiheng Lyu, Kevin Blin, Fernando Gonzalez, Max Kleiman-Weiner, Mrinmaya Sachan, Bernhard Schölkopf.

Analyzing the Role of Semantic Representations in the Era of Large Language Models. *(ARR 2023 acceptance; commited to NAACL 2024)*
Zhijing Jin*, **Yuen Chen***, Fernando Gonzalez Adauto*, Jiarui Liu, Jiayi Zhang, Julian Michael, Bernhard Schölkopf, Mona Diab.

CausalCite: A Causal Formulation of Paper Citations. *(ARR 2024 Under Review)*.
Ishan Kumar Agrawal*, Zhijing Jin*, Ehsan Mokhtarian, Siyuan Guo, **Yuen Chen**, Negar Kiyavash, Mrinmaya Sachan, Bernhard Schölkopf.

*Equal Contribution

## HIGHLIGHTED PROJECTS

**Time Series Analysis on Semiconductor Processing Tools** Fall 2022

- Developed a machine learning model to predict the performance of wafer production process
- Condensed 560k+ data into 216 data by Fast Fourier Transform, sinusoidal regression, and quadratic regression
- Achieved 87% prediction accuracy on wafer metrology with a linear regression model

**Offline Evaluation of Bandit Algorithms** Spring 2022

- Evaluated bandit algorithms with importance-weighted losses on "R6B Yahoo! Front Page Today Module User Click Log Dataset"
- Investigated the performance of UCB1, EXP3, and random strategy compared to the theoretical performance bound
- Achieved 2.6x better performance on EXP3 algorithm than the theoretical performance lower bound

**Medical Images Segmentation** Spring 2022

- Implemented U-Net with PyTorch to segment the blood vessels on photographs of the retina
- Trained the neural network with sample-splitting and output segmented images for the test image data
- Achieved <10% binary cross-entropy loss after 40 training epochs

**Representation Learning and Generative Modelling on MNIST Dataset** Spring 2022

- Performed dimensionality reduction on MINIST dataset using PCA, Autoencoder (AE), and Variational Autoencoder (VAE)
- Optimized binary cross-entropy loss of AE from 0.005 to 0.002 and evidence lower bound loss of VAE from 500+ to 200
- Synthesized new image data by sampling from the latent space of VAE

**R Packages Text Analysis** Spring 2021

- Performed frequency text analysis by regular expression and visualized the outcome by R data visualization tool
- Built a web application with R-shiny that demos the analysis of a distribution of R packages' title length

**Modified Brent's Method** Spring 2021

- Implemented the root finding algorithm proposed by Wilkens and Gu in MATLAB
- Performed bisection method, inverse quadratic interpolation, and secant method to reduce the numbers of function evaluations
- Achieved 92% test case coverage within tolerance of 1e -15 in terms of efficiency

## PROFESSIONAL & LEADERSHIP EXPERIENCE

**Statistics Course Reader| University of California, Berkeley** September 2022 - December 2022

- Assisted and graded 800+ students on homework, labs, exams weekly in Stat 20: Introduction to Statistics
- Co-managed 8 sections of lectures with Prof. Andrew Bray and other course staff on course content

**Math Tutor | Irvine Valley College** August 2019 - May 2022

- Conducted 1-on-6 weekly sections and hosted office hours for 200+ students weekly
- Designed weekly problem sets for teaching uses in fields of trigonometry, calculus, and differential equations
- Earned high satisfaction in teaching from students in Fall 2021 and Spring 2022

**Commissioner | Student Government of Irvine Valley College** August 2018 - September 2019

- Coordinated with 6 other committees on planning and advertising campus events
- Allocated $800,000 in funds to student organizations, intercurricular programs, and scholarships