Yuen Chen

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

EDUCATION

University of Illinois at Urbana-Champaign — Expected May 2028
Ph.D. Candidate in Computer Science, advised by Prof. Hari Sundaram and Prof. Han Zhao
Research Topics: Causal Machine Learning, Trustworthy Machine Learning

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, Domain Generalization
[Programming Languages] Python, R/R Studio, MATLAB
[Machine Learning Frameworks and Libraries] PyTorch, TensorFlow, NumPy, Pandas, DoWhy

RESEARCH EXPERIENCE

Applied Research Intern | AI Foundation, Capital One — June 2025 - August 2025
Mentor and Manager: Rizal Fathony & Nam Nguyen
- Designed a novel graph and sequence architecture for customer behavior modeling

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

CausalDetox: Causal Head Selection and Intervention for Language Model Detoxification. *(ACL 2026)*
Yian Wang, **Yuen Chen**, Agam Goyal, Hari Sundaram

Moment Alignment: Unifying Gradient and Hessian Matching for Domain Generalization. (UAI 2025)
Yuen Chen, Haozhi Si, Guojun Zhang, Han Zhao

Breaking Bad Tokens: Detoxification of LLMs Using Sparse Autoencoders. (EMNLP 2025)
Agam Goyal, Vedant Rathi, William Yeh, Yian Wang, **Yuen Chen**, Hari Sundaram

DIVEBATCH: Accelerating Model Training Through Gradient-Diversity Aware Batch Size Adaptation. *(arXiv 2025)*
Yuen Chen, Yian Wang, Hari Sundaram

Bridging the Divide: End-to-End Sequence-Graph Learning. *(arXiv 2025)*
Yuen Chen, Yulun Wu, Samuel Sharpe, Igor Melnyk, Nam H. Nguyen, Furong Huang, C. Bayan Bruss, Rizal Fathony

Causally Testing Gender Bias in LLMs: A Case Study on Occupational Bias. *(Findings of NAACL 2025)*

Yuen Chen*, Vethavikashini Chithrra Raghuram*, Justus Mattern*, Rada Mihalcea, Zhijing Jin.

Analyzing the Role of Semantic Representations in the Era of Large Language Models. *(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. *(Findings of ACL 2024)*

Ishan Kumar Agrawal*, Zhijing Jin*, Ehsan Mokhtarian, Siyuan Guo, **Yuen Chen**, Negar Kiyavash, Mrinmaya Sachan, Bernhard Schölkopf.

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.

CauSciBench: A Comprehensive Benchmark on End-to-End Causal Inference for Scientific Research. *(Preprint)*

Sawal Acharya, Terry Jingchen Zhang, Andrew Kim, Anahita Haghighat, Sun Xianlin, Pepijn Cobben, Rahul Babu Shrestha, Maximilian Mordig, Jacob T Emmerson, Furkan Danisman, **Yuen Chen**, Clijo Jose, Andrei Ioan Muresanu, Justin Cui, Jiarui Liu, Yahang Qi, Punya Syon Pandey, Yinya Huang, Bernhard Schölkopf, Mrinmaya Sachan, Zhijing Jin.

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

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

*Equal Contribution