

## Education

- 2024–Present **PhD, Statistics**, *Columbia University*, New York, NY.
- 2022–2024 **MS, Computer Science**, *Columbia University*, New York, NY, GPA: 3.99/4.00.
- 2017–2021 **BS, Computer Science (honors), Mathematical Economics, Summa Cum Laude**, *University of Richmond*, Richmond, VA, GPA: 3.94/4.00.
- Fall 2019 **Visiting Student**, *University of Edinburgh*, Edinburgh, Scotland, GPA: 4.0/4.0 (converted).  
**Course Highlights:** Machine Learning and Pattern Recognition, Accelerated Natural Language Processing, Fundamentals of Operational Research

## Peer Reviewed Publications

- ML4LMS Workshop @ ICML 2024 S. Chang, **Z. Chen**, B. Dumitrascu, D. A. Knowles, *CellFlows: Inferring Splicing Kinetics from Latent and Mechanistic Cellular Dynamics* in ICML'24 Workshop ML for Life and Material Science: From Theory to Industry Applications. 2024.
- EMNLP 2022 **Z. Chen** D. Verdi do Amarante, J. Donaldson, Y. Jo and J. Park, *Argument Mining for Review Helpfulness Prediction* in Proceedings of the Empirical Methods on Natural Language Processing (EMNLP). 2022.
- DSAA 2021 **Z. Chen** and J. Park, *Analyzing Cultural Assimilation through the Lens of Yelp Restaurant Reviews* at the 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2021).

## Research Experience

- 2022–2024 **Graduate Research Assistant**, *Department of Computer Science: Dr. David Knowles*, Columbia University, New York Genome Center.
- Developing systems to improve experimental design of massively parallel reporter assays (MPRAs) in order to better study the sequence determinants of alternative splicing
  - Utilizing deep generative models, convolutional neural networks, and conditional neural processes in order to more efficiently learn splicing outcomes from DNA sequences
  - Training DNA languages models with ATAC-Seq annotations in order to understand variant effects on chromatin accessibility
- 2020–2021 **Honors Thesis**, *Department of Computer Science: Dr. Joonsuk Park*, University of Richmond.
- Developed methods for graph-based argument mining using large language models, transfer learning, and dependency parsing methods
  - Utilized annotations from argument mining model to improve performance on helpfulness prediction of Amazon product reviews
  - Implemented methods in PyTorch and Huggingface's transformers library
  - Project culminated into a EMNLP 2022 publication
- Spring, Summer 2020 **Research Assistant**, *Department of Computer Science: Dr. Joonsuk Park*, University of Richmond.
- Developed probabilistic graphical models to perform topic modeling on online review data for cultural analysis
  - Implemented graphical models with collapsed Gibbs Sampling in Python and Cython
  - Project culminated into a DSAA 2021 publication
- Summer 2019 **Summer Research Fellowship**, *Department of Computer Science: Dr. Prateek Bhakta*, University of Richmond.
- Developed algorithms towards fair and robust liver donation allocation strategies, focusing on maximum-flow algorithms and their properties
  - Programmed simulations in Python and C++ to test developed algorithms
  - Scrapped and cleaned medical and geographical data from publicly available sources for use in simulations
- 2018–2019 **Research Assistant**, *Department of Psychology: Dr. Cindy Bukach's Lab*, University of Richmond.
- Programmed experiments using Neurobehavioral Systems Presentation, Codex Superlab, and Python which investigated the Other Race Effect (ORE) and the Angry Black Bias
  - Performed data cleaning and analysis using Python and R
  - Created interactive web simulations using HYPE

## Work Experience/Service

- Summer 2021–Present **Machine Learning Software Engineer**, *CarMax*.
- Building and maintaining personalized recommendations systems as well as customer segmentation systems for search engine optimization
  - Building microservices and common Python libraries in order to support data sourcing and model orchestration
  - Building CI/CD pipelines in Azure DevOps in order to deploy cloud resources to Azure
  - Training, deploying, and serving model predictions through AzureML, Apache Spark, and Databricks

- Fall 2023 **Reviewer**, *Nature Machine Intelligence*.
- Spring 2019, **Grading Assistant**, *Department of Computer Science*, University of Richmond.
- Spring 2021 Graded weekly homeworks and programming assignments for two sections of Data Structures, taught by Dr. Jory Denny, as well as one section of Computer Organization, taught by Dr. Prateek Bhakta
- Spring 2019 **Grading Assistant**, *Department of Mathematics*, University of Richmond.  
Graded weekly homeworks for two sections of Linear Algebra, taught by Dr. William Ross

---

## Technical Skills

- Relevant Coursework Foundations of Graphical Models, Machine Learning in Genomics, Deep and Generalized Probabilistic Models, Neural Networks and Deep Learning, Geometric Data Analysis, Interpretable Machine Learning for Inverse Problems in Biology
- Software PyTorch, TensorFlow, Python,  $\text{\LaTeX}$ , Apache Spark, C++, Microsoft Azure Cloud Services, Google Cloud Platform (GCP), Azure Cosmos DB, Azure DevOps, Azure ML

---

## Honors

- 2021 **Phi Beta Kappa**, University of Richmond
- 2017–2021 **University of Richmond Science Scholar Award**: Full tuition merit scholarship for 4 years
- 2017–2021 **National Merit Scholarship**
- 2017 **National AP Scholar**
- 2016 **Most Interesting Hack** at Vassar College Hackathon