# Haochong (Rogers) Yang

(437) 981-3043 haochong.yang@mail.utoronto.ca HaochongYANG haochongyang.github.io LinkedIn

## Education

**Honours Bachelor of Science in Statistics and Computer Science** **Sep 2021 – Present**
*University of Toronto (U of T) | Toronto, ON, Canada* *Expected in May 2025*

- *cGPA: 3.97/4.00*
- *Relevant Courses: Linear Algebra, Calculus, Probability, Data Analysis, Data Structure, Mathematical Statistics, Database, Time Series, Multivariate Statistics, Machine Learning, Neural Network, Stochastic Processes*

## Publication

**Yang, H.**; Huang, M.; Chen, X.; He, Z.; Pu, S. Enhanced Real-Life Data Modeling with the Modified Burr III Odds Ratio–G Distribution. *Axioms* 2024, 13, 401. https://doi.org/10.3390/axioms13060401 Published

**Yang, H.**; Sun, Y. H.; Lee, K. A Novel Approach for Efficient Multi-class Anxiety Level Prediction Approximation for Long Assessments. Submitted

## Awards

Summer Undergraduate Data Science Scholarship — Data Science Institute | 2024
C.L. Burton Scholarship for Mathematics and Physical Sciences — University College | 2023
University of Toronto Scholar Award — University of Toronto | 2022
Louis Savlov (UC'37) Scholarships in Sciences and Humanities — University College | 2022
Dean's List Awards — Faculty of Arts and Science | 2022, 2023, 2024

## Technical Competencies

**Languages** Python, Java, JavaScript, Shell
**Data** R, SQL, Tableau, PowerBI, Kafka
**Frameworks** TensorFlow, PyTorch, Langchain, Django
**Packages** Pandas, NumPy, Matplotlib, Scikit-Learn

## Research Experience

**Heterogeneity of Single Cell DNA Structure Analysis**
*Department of Statistical Sciences | U of T* *Supervisor: Prof. Elena Tuzhilina*

- Conducted individual research under professor's guidance, with regular progress updates via weekly meetings
- Explored the quantitative methods such as local contacts difference and cosine distance for classifying single cells based on DNA Hi-C data and contact matrices
- Applied dimension reduction techniques such as PCA, t-SNE, and Uniform Manifold Approximation & Projections through proofs and coding to find representations of cells in low dimensions for clustering

**Efficient Anxiety Level Prediction using Machine Learning**
*Department of Applied Psychology and Human Development | U of T* *Supervisor: Prof. Kang Lee*

- Enhanced machine learning model for DASS-42 mental health survey result prediction, reducing the length of the original assessment while maintaining high prediction accuracy
- Developed an ensemble model that achieved an accuracy of 85% in classifying anxiety levels across five categories with high reliability, while baseline chance level is 20% in five-level classification tasks
- Implemented a web app to enable anxiety assessments based on the trained model using React and Heroku for cloud deployment.
- First-authored a manuscript detailing the research method and findings and submitted to the *Journal of Anxiety Disorders*

**Design LLM Systems in Mental Health Assessments**
*Department of Applied Psychology and Human Development | U of T* *Supervisor: Prof. Kang Lee*

- Developed a system of LLMs to approximate the role of psychologists to conduct mental health assessments
- Led prompt engineering work for LLMs and coded a duo-agents psychological analysis robust system
- Worked on front and back-end to enable the multimodality detection of facial expression, tone, and actions
- Collaborated with team members to create research proposals, draft ethics reports, and plan clinical studies

**Social Theories Validation using Reinforcement Learning (RL)**
*Department of Psychology | U of T* *Supervisor: Prof. William Cunningham*

- Leaded Leaky Emotions project to research on how agent emotions affect others' behavior in a social setting
- Fine-tuned RL environment to validate various social cognition theories with support of neural networks
- Designed visuals for training process, facilitating rigorous game logic execution and components modification
- Leveraged ML techniques to train CPU agents for optimal game results, and performed statistical analysis on game data to show the agents gaining social impacts on others to receive higher mutual rewards

## Industry Experience

**Data Analyst (Co-op)** **May 2023 – Aug 2023**
*Toromont CAT | Concord, ON, Canada*

- Trained a large language model with documentation and system data using LangChain and OpenAI API to develop a company chatbot "CatGPT" which will benefit over 2000 people as users to look for company business solution
- Optimized the Strategic Asset Management platform through the implementation of Power BI and Python, enabling real-time monitoring and efficient tracking of heavy mining machinery status and operational activities
- Created predictive model with Matplotlib to keep track of machine usage patterns and forecast component replacement
- Collaborated with 16 Canadian mine owners to update machine component status, resulting in a significant reduction of $57 million in overdue value

**Software Developer (Internship)** **May 2022 – Aug 2022**
*BL Innovare | Markham, ON, Canada*

- Spearheaded daily data analysis, software development, and support operations for a leading company specializing in vehicle inspection and maintenance products and services
- Enhanced the performance of Bodyguard 2.0, an advanced machine learning application designed to simulate human perceptual and decision-making processes for vehicle inspection and damage analysis
- Contributed to the back-end development of a language translating platform using Python and the Django framework
- Demonstrated proficiency in crafting intricate SQL queries to process customer data and conduct in-depth data analysis