

Gabriel Cha

+1 (832) 993-4499 | gsc2133@columbia.edu | linkedin.com/in/gabrielcha | **US Citizen**

Education

Columbia University

New York, NY

M.S. in Data Science

Sept 2025 - Dec 2026

- Specialization in Deep Learning for Computer Vision, Speech, and Language

University of California, San Diego

San Diego, CA

B.S. in Data Science

Sept 2021 - March 2025

- 3.75 GPA: 9x Provost Honors Recipient, Halicioğlu Data Science Institute Research Scholarship (\$6500)

Experience

DSTL Lab

San Diego, CA

Research Assistant

Oct 2024 - Present

- Developed a full-stack chatbot powered by LLMs and integrated into JupyterLab, written in Python and TypeScript, and engineered to assist instructors in generating lecture content; scheduled for deployment in the upcoming year
- Collaborated with instructors (Professor Sam Lau) and subject matter experts to fine-tune the chatbot's natural language processing capabilities to ensure context-aware responses tailored to instructor's needs

Edison International

Rosemead, CA

Data Analytics Intern

June 2024 - Sept 2024

- Built and deployed data ingestion and validation pipelines to process millions of 5-minute solar energy data; the system flagged out-of-bound metrics on monthly basis and reduced manual validation time from 2 days to 5 minutes
- Developed ETL pipeline and internal interface written in python to automate third-party data requests; reduced retrieval time from 4 days to 10 minutes; successfully serviced requested data on the first day of deployment!

UC San Diego

San Diego, CA

Teaching Assistant

Sept 2022 - March 2025

- Worked closely with faculty and academic departments to teach advanced concepts in algorithms, statistics, and machine learning in Python to over 1,200 students across 9 quarters; consistently earning 100% recommendation
- Designed and deployed 2 programming assignments with professors: bus route optimization with graph algorithms and loan analysis using statistical tests; built custom autograding scripts and completed by 500+ students

Projects

Concept Bottlenecks for LLMs

Dec 2024

- Developed platform for integrating Concept Bottleneck Layers (CBLs) into LLMs to visualize top-k neuron activations, prune biased concepts, and increase trustworthiness of traditionally black box models; inspired by Weng et al. (2024)

Are you sad? Speech Emotion Recognition

Sept 2023

- Led a \$6,500 UCSD-funded research project to classify emotions in speech, performed data extractions to visualize frequencies, built pipelines to add augmentations, and achieved 86% accuracy using convolutional neural networks

Gatsby-Inspired Language Model

Feb 2023

- Built N-Gram language model leveraging web-scrape text from The Great Gatsby to uncover linguistic patterns, predict word probabilities, and generate statistically-driven, Gatsby-inspired text!

Skills

Languages	Python, Java, Javascript, Typescript, SQL
Artificial Intelligence/ML	TensorFlow, PyTorch, Hugging Face, Scikit-learn, NumPy, Pandas, OpenCV
Data Analytics	Pandas, Polars, Apache Spark, Excel, Plotly, Tableau, Regex, BeautifulSoup
Web Development	React, HTML, CSS, Node.js, REST API, Docker, Flask, Django
Software Tools	GitHub, Docker, Bash, Jupyter Notebook, VS Code