

Christopher Cha

☎ (805) 443-3829 | ✉ ccha@ucsd.edu | 🏠 ccha.site | 🌐 Rinsworth | 📄 christopher-cha

Education

University of California San Diego Jacobs School of Engineering

B.S. IN COMPUTER SCIENCE

M.S. IN COMPUTER SCIENCE, FOCUS IN ARTIFICIAL INTELLIGENCE

3.79 BS GPA, 4.00 MS GPA

BS: Jun. 2019 - Jun. 2022

MS: Jun. 2022 - Dec. 2023

Experience

Machine Learning Engineer/Software Engineer (Co-op)

EMD ELECTRONICS (MERCK KGAA)

Darmstadt, Germany/Carlsbad, CA

Jun. 2022 - Present

- Focusing on applying and deploying machine learning models to semiconductor chemistry research.
- Used **AWS EC2** and **S3** to train a variational autoencoder to map semiconductor precursors to a latent space.
- Applied a sparse dataset of less than 1,000 molecules to a latent space and an 85% accuracy of mapping a chemical to the space and getting back the original molecule, as well as a 99% validity of the entire latent space.
- Created web interfaces of **Pytorch** and **SKlearn** models using **React.js** and **Node.js**.
- Optimized existing chemical similarity code bases by reducing I/O usage by 90% and improving runtime by 25%.

Natural Language Processing Researcher

SAN DIEGO SUPERCOMPUTER CENTER

La Jolla, CA

Mar. 2022 - Present

- Currently assuming responsibility over a relation extraction project related to biology papers.
- Used **Kubernetes** and **Docker** to schedule and run Python scripts.
- Performed relation extraction on the DrugProt and ChemProt datasets using **Pytorch**.
- Began compatibility migration to run model training and inference on Intel Habana Goya and Gaudi AI accelerators.

Academic Research

Research/Development Intern

THE DESIGN LAB

La Jolla, CA

Jun. 2021 - Feb. 2022

- Created an idea dataset collection system using **GraphQL** and **Python**.
- Processed metadata text using **Spacy** and **NLTK** using **AWS EC2** and created a dataset (11,000+ unique items) using **Pandas** and **MongoDB**.
- Implemented a semantic metadata comparison using SentenceTransformers (**BERT**) in order to return related ideas.
- Developed and hosted the metadata comparison model using **Flask** and **Nginx**.

Natural Language Processing Undergraduate Researcher

UC SAN DIEGO CSE DEPARTMENT/SCRIPPS INSTITUTION OF OCEANOGRAPHY

La Jolla, CA

Dec. 2020 - Jan. 2022

- Cleaned and labeled 161 national climate action plans using **BeautifulSoup** in Python.
- Applied Latent Dirichlet allocation and TF-IDF to a number of country characteristics.
- Trained language models (**BERT**) in **PyTorch** to create labels and skill metrics.

Publications

Tom Corringham, Daniel Spokoyny, Eric Xiao, **Christopher Cha**, Colin Lemarchand, Mandeep Syal, Ethan Olson, Alexander Gershunov (2021). "BERT Classification of Paris Agreement Climate Action Plans." *Tackling Climate Change with Machine Learning Workshop at ICML 2021*.

Xiaotong (Tone) Xu, Khuyen Le, **Christopher Cha**, Samuel Fleet, Steven P. Dow (2021). "Crowdsourcing Exploratory Cues for Idea Browsing and Inspiration Discovery." *HCOMP 2021*.

Projects/Coursework

Fast Text Review Spoiler Detection Model

Feb. 2022

- Trained a DistilBERT classifier of text spoilers using **Pytorch** with an accuracy of 83% on the test set without any other context.
- Compared various models such as **LSTMs**, **Transformers** (RoBERTa), and other classical models.
- Deployed the model using **Flask** and **MongoDB** with a focus on performance and latency.

Ratings Prediction on a Crowdsourced Manga Dataset

Sep. 2021 - Dec. 2021

- Curated a dataset from a social cataloging anime/manga database using **GraphQL**.
- Cleaned and processed data using **Pandas**, **Scikit learn**, and **Scipy**.
- Created a feature vector from 13 features within our dataset to describe the content.
- Achieved a **136 test MSE** using a K nearest neighbor algorithm (K=50) on our feature vector, compared to our trivial baseline of 169 MSE.

Leadership/Activities

ACM@UCSD Technical Event Director

ASSOCIATION FOR COMPUTING MACHINERY

La Jolla, CA

Jun. 2020 - Present

- Organized and present workshops to 100+ members for 2020-2021 school year
- Created workshops focusing on programming languages (APL and \LaTeX) and natural language processing.
- Partnered with Workshops and Info Sessions for Engineers in order to teach programming languages to engineers at UC San Diego.
- Managed a shadowing program with 10+ members to teach presentation skills and create workshops.

Skills

Languages	Python, Java, C, C++, Javascript (Node.js), ARMv6 Assembly, Bash, R, Matlab
Tools/DevOps	Git, Unix, Jupyter, Kubernetes, MongoDB, GDB/Valgrind, Flask, AWS, React, Docker, SQL, Node.js, jQuery
Packages	Pandas, Numpy, PyTorch, Keras, Sklearn, Scipy, Spacy, NLTK, Matplotlib, Seaborn, Streamlit
Coursework	Software Engineering (Agile), Deep Learning, Data Science, Recommender Systems, Computer Security, Operating Systems