

# Yong Nan Chang

[eclec.chang@gmail.com](mailto:eclec.chang@gmail.com) | 208-346-2531 | [linkedin.com/in/yong-nan-chang-132875192](https://www.linkedin.com/in/yong-nan-chang-132875192)

---

## WORK EXPERIENCES

### Data Scientist | UC Berkeley, Haas School of Business | Remote | Aug 2023 - Present

- Improved accuracy from **62% to 71%** for a **text classification** model with **NLP Transformer** and **Transformer summarization**
- Created a name-matching model with **TF-IDF**, **Sentence Transformer**, **BERT Embedding**, and **Fuzz** technique.
- Hosted a **local LLM model** for long-token text classification tasks and **fine tune GPT3.5** and **LLAMA3** with QLora

### Data Science Intern | CrowdStrike | Berkeley, CA | Aug 2023 - May 2023

- Developed a **GCP BigQuery database** for CVE data, integrating open-source data and improving data completeness.
- Hosted and built **LLM text classification model** with **llama2** and **prompt engineering** to classify CVE data

### Software Development Engineer in Test | Stampin' Up! | Salt Lake City, UT | May 2022 - Aug 2023

- Designed and implemented **ETL pipeline** architectures that extract JSON data from a **REST API** with **OAuth** authentication and transform it using **Apache Spark** in **Azure Synapse Analytics** and **Azure Data Lake Storage**.
- Built **E2E automation testing** with **Cypress** and **Postman** to test the front-end user interface and backend API endpoint.
- Improved processing time by 50% by creating automation scripts with parallelization in **C#**, **Bash**, and **SQL** for data validation.
- Created automated ETL tests using **Python**, **Great Expectations**, and **Faker** to generate test scenarios that verify data accuracy.
- Conducted and executed a backend production launch **test plan**, and created mock data to meet all testing requirements.

### Data Scientist | Contract | Remote | Nov 2022 - Aug 2023

- Built and trained a **91% accuracy** semantic segmentation model using **FCN-ResNet** for cancer detection in **AWS SageMaker**.
- Designed and implemented custom **REST API endpoints** for customers to upload pictures for prediction using **AWS Lambda**, **API Gateway**, **S3**, and **SageMaker**.

### Business Intelligence Engineer | Yatta Ops | Provo, UT | Dec 2021- Apr 2022

- Optimized the **data warehouse** pipeline loading time by 20% by removing duplicate rows and reducing unused functions.
- Automated ETL Pipeline with **Python** to extract and load sales data from PDF reports into an **MS SQL server**
- Monitored the data warehouse pipeline by optimizing job schedules and troubleshooting loading errors in **Domo Workbench** and **SQL server** to ensure consistency and accuracy in the **dashboard**.
- Improved the **loan default rate prediction** by removing outliers and inaccurate data points for the **XGBoost ML model**.
- Built **ETL pipelines** to extract data from **REST API** endpoints and update calculated predictions on the Domo dashboard.

### Business Intelligence Data Analyst Intern | Frazil | Salt Lake City, UT | May 2021 – Aug 2021

- Boosted model accuracy by 12% for a sales performance forecasting model using **exponential smoothing**.
- Enhanced the **XGBoost ML model** by 7% accuracy with feature engineering that predicts profitability with **AWS SageMaker**.
- Developed a **product recommendation algorithm** for retailer selling strategies using **ANOVA**, **t-test**, and **Beta regression**.
- Cleaned and retrieved **HeidiSQL** data to unify a **data warehouse**.

---

## RESEARCH EXPERIENCES

### Research Assistant | Brigham Young University | Provo, UT | Jun 2022 – Jan 2023

- Analyzed student video-watching behavior on YouTube using **multiple linear regression** and **gradient descent models** to find the correlation between learning outcomes and actions to improve student grades.
- Cleaned data on a time series dataset using **data mining**, **outlier detection**, and **sequence pattern mining** techniques to remove inaccurate data and outliers.

---

## EDUCATION

### University of California Berkeley, May 2024

GPA 4.00

Master of Engineering in Industrial Engineering & Operations Research

### Brigham Young University, Dec 2022

Bachelor of Science in Statistics, Data Science Emphasis, Information Technology Minor

GPA 3.99

Full Scholarship, Magna Cum Laude, Dean's list, Mu Sigma Rho Society

---

## SKILLS

**PROGRAMMING LANGUAGES:** Python, R, C#, .NET, C++, SQL, Bash, Linux

**DATA ENGINEERING:** Data Modeling, MS SQL Server, Apache Spark, ETL Pipeline, Data Lake/Warehouse, Distributed Computing, JSON

**DATA SCIENCE:** Machine Learning, Deep Learning, Time Series Analysis, Linear Regression, Tableau, Pandas, Pytorch, scikit-learn, NLP

**CLOUD AND SOFTWARE SERVICES:** Azure (Synapse Analytic, Databrick, Pipeline), AWS (Sagemaker, Lambda, S3, EC2, API Gateway, Cognito) Git, Docker, Unit and E2E Testing, iCEDQ ETL Testing, Great Expectation, Cypress Automation Testing, DOMO, MongoDB, Excel

**CERTIFICATION:** AWS Certified Cloud Practitioner, Tableau Desktop Specialist