# Sean (Shuo) Guan

Tel: +1 (646) 403-7894 Email: seanguan325@gmail.com

## Work Experience

**Credit Suisse** Eleven Madison Avenue, New York, NY, 10010; Feb.2022-Now

*Position: Technical Analyst (Full-time);*

- Worked as the data/ML analyst and the full stack developer in the team;
- Managed to designed and built multiple machine learning interpreters of the black-box models for 3 projects;
- Built and deployed the backend (Hive tables, spark logic, hdfs, etc.) and the REST API for multiple projects;
- Built and maintained a web App for a data exception remediation system;

**Credit Suisse** Eleven Madison Avenue, New York, NY, 10010; June 2021-Aug 2021

*Position: Summer Intern;*

- Lead the intern group project and achieved outstanding results on both group project and independent project;
- Designed the machine learning interpreter for the isolation forest model of an CFO anomaly detector;
- Designed and built a two-phase sequence tagger for user-generated texts, using customized deep learning model.

**Dell EMC Inc.** Building 6, Chuangzhitiandi, 433 Songhu Road, Shanghai, China; July.2018

*Undergraduate Summer Internship Program;*

- Implemented and maintained a blockchain platform and an interacting system based on Hyperledger Fabric;
- Had a short research on the outlier detection based on CNN and LSTM.

## Education

**M.S.: New York University** (Courant Institute of Mathematical Sciences) New York, Sep.2019 – Jan.2022

Major: **Computer Science** **GPA: 3.84/4.0**

**B.S.: Tongji University** Shanghai, China, Sep.2015 – Jun.2019

Major: **Computer Science and Technology**

Minor: **Applied Mathematics** (Sep.2015 -Feb.2017)

**GPA: 4.41/5.0 Major GPA: 4.71/5.0**; Honors: Shanghai Scholarship (2/112), Outstanding Graduate of Shanghai (6/112)

**Language Skills:** Python, C, C++, SQL, Java, MATLAB, C#.Net

**Tools & Framework Skills:** TensorFlow, PyTorch, H2O.ai, Git, Spark, HDFS, Hive, Impala, Shell, Docker, k8s, Jenkins

## Projects & Research

### Work Projects:

**Operation Risk and Controls Anomaly Detector** Aug.2022 - Now

- Built, deployed and maintained the REST API of the ML model interpreter and the user access control;
- Designed and implemented the ML interpreter for the isolation forest model in the anomaly detector;
- Built and maintained backend HDFS storage and part of the database schemas for the system;

**Smart Data Exception Remediation System (Professor-Ex)** Apr.2022 - Now

- Built and maintained the web App for monitoring the data remediation model performance of the system and providing the visualized interpretation for the remediation model, using Dash, html and css;
- Built the machine learning interpreter framework for the random forest and the GBM models used in the system;
- Managed to complete the production parallel release for the project;

**No-Code Automotive AI Platform (Drebin)** Mar.2022 - Oct.2022

- Built and maintained the web App for data visualization and model control of the platform using Dash, html and css;
- Built and optimized the platform backend system and achieved a faster responding speed;
- Explored and built the interpretation method for semantic analysis and text summarization/generation;
- Made the ML models (anomaly detection, clustering, semantic analysis) in the platform interpretable;
- Designed a new visualizable ML interpreter for K-means clustering model

**Product Control Anomaly Detector** Feb.2022 - May.2022

- Built the web App for the system using Dash, html and css;
- Optimized and maintained the backend functionality, maintained its HDFS storage and database schemas

**User Texts Semantic Tagging Intern Project** Jun.2021 – Aug.2021

- Designed a new character level CNN structure for the user generated text scenario, achieved a high accuracy (95%);
- Built a 2-phase sequence tagger, including a phrase extractor based on part-of-speech tagging and a phrase classifier based on deep learning;
- Built a prediction API for the showing the functionality in the frontend web App

### Research Projects:

**Extract, Select and Rewrite: A New Modular Summarization Method** Jan.2021 - Jan.2022

*Advisor: Assistant Prof. He He; Research project in NYU Courant*

- Proposed a three-phase (knowledge extraction, content selection and rewriting) modular abstractive summarization method based on the knowledge triples;
- Designed multiple subtask datasets for training the modules in the framework;
- Fine-tuned multiple pre-trained language models on the sub datasets for the content selector and rewriter modules;
- Showed SOTA-competitive performance on multiple datasets and good modularity;

**Knowledge and Keywords Augmented Abstractive Sentence Summarization** Sep.2020 - Jan.2021

*Advisor: Prof. Zhihua Wei; Research project with Tongji University*

- Proposed a novel short text abstractive summarization augmented by knowledge and topic keywords;
- Constructed a special linearized knowledge structure for short texts;
- Designed the tri-encoders structure, tri-copy mechanism and hierarchical attention for the language model;
- SOTA-competitive performances on multi-language summarization datasets.

**A New Hybrid Model for Recommendation System based on DNN and RBM** Mar.2019 - June.2019

*Advisor: Prof. Jiujun Cheng; Bachelor Graduation Thesis*

- Established and optimized a content-based 7-layer DNN (Text CNN embedded) structure and feature engineering;
- Designed and implemented a RBM collaborative filtering model, and tried to build DBN model for recommend;
- Used 3 methods to combine two models and proved that the hybrid model had a much better result than others.

**An Improved LeNet-5 Gas Identification Structure for Electronic Noses** Mar.2018 - Sep.2018

*Advisor: Prof. Guangfen Wei; Sponsored by project of Natural Science Foundation of China (NSFC) (No. 61174007)*

- Designed a data augmentation method for expanding the gas identification dataset;
- Built and tested an improved convolutional neural network structure based on LeNet-5 for gas identification of electronic noses;
- Got the result that the final gas identification accuracy rate reaches 99.67% with the optimized structure.

## Publications

1. **Shuo Guan**, Vishakh Padmakumar, He He. "Extract, Select and Rewrite: A New Modular Summarization Method" Manuscript (2022), had an oral presentation on MASC-SLL 2022.

2. **Shuo Guan**, Ping Zhu, Zhihua Wei. "Knowledge and Keywords Augmented Abstractive Sentence Summarization." *EMNLP 2021 Workshop on New Frontiers in Summarization* (2021): 25-32.

3. Guangfen Wei, Gang Li, **Shuo Guan**, Jie Zhao, Xue Sun. "Study on an Improved LeNet-5 Gas Identification Structure for Electronic Noses." *2018 IEEE SENSORS* (2018): 1-4.

4. **Shuo Guan**. "Analysis of Optimal Pricing Model of Crowdsourcing Platform Based on Cluster and Proportional Sharing." *2018 6th International Symposium on Computational and Business Intelligence (ISCBI)* (2018): 99-103.