# Shuyang Li

shuyangli94@gmail.com ▪ shuyangli.me ▪ github.com/shuyangli94 ▪ (314)-660-6660
Office #4146, 9500 Gilman Dr, San Diego, CA 92122

## EDUCATION

**University of California, San Diego**, San Diego, CA (PhD) *Exp.* 2023
Advisor: Dr. Julian McAuley
Area: Computer Science and Engineering, Machine Learning

**Princeton University**, Princeton, NJ (BSE) June 2016
Major: Operations Research and Financial Engineering GPA: 3.71
Certificates: Applications of Computing; Statistics and Machine Learning
Thesis: *Exploring Rich Features for Sentiment Analysis with Various Machine Learning Models*

## PUBLICATIONS

**Li, S.** and McAuley, J. (2020). Recipes for Success: Data Science in the Home Kitchen. *Harvard Data Science Review.*

Mao, H., **Li, S.**, McAuley, J., Cottrell, G. (2020). Speech Recognition and Multi-Speaker Diarization of Long Conversations. *INTERSPEECH*

Majumder, B.*, **Li, S.***, Ni, J., McAuley, J. (2020). Interview: A Large-Scale Open-Source Corpus of Media Dialog. *ArXiv Preprint*

Majumder, B.*, **Li, S.***, Ni, J., McAuley, J. (2019). Generating Personalized Recipes from Historical User Preferences. *Empirical Methods in Natural Language Processing.*

**Li, S.***, Majumder, B.*, McAuley, J. (2019). Cooking Common Sense: Personalized Recipe 'Tweak' Inference via Common Sense Reasoning. *SoCal NLP Symposium.*

Huang, D., **Li, S**., Dhaka, A., Story, G.M. and Cao, Y.Q. (2012). Expression of the transient receptor potential channels TRPV1, TRPA1 and TRPM8 in mouse trigeminal primary afferent neurons innervating the dura. *Molecular Pain*, 8 (1), 66–85.

## HONORS AND AWARDS

*Finalist*, Qualcomm Innovation Fellowship 2020-2021
"Toward Personalized and Multimodal Conversational Recommender Systems"

*Finalist*, Amazon Alexa Prize SocialBot Grand Challenge 2019 (UCSD Team) 2019-2020

## RESEARCH AND WORK EXPERIENCE

**University of California San Diego** September 2018-Present
*Graduate Student Researcher*

- Researching and developing models to incorporate knowledge, subjectivity, and personalization in dialogue/language modeling and generation at the intersection of NLP and recommender systems
- Researcher, architect, and lead developer for the UCSD Amazon Alexa Prize Socialbot Grand Challenge team

**Google, Kaggle Datasets (Google Cloud AI)** June 2019-September 2019
*Software Engineering Intern (PhD)*

- Built a framework for automatically generating semantic tags for datasets based on free-text metadata
- Implemented metrics for dataset discoverability and search success
- Doubled size of tag ontology and tripled tag coverage across all public datasets on Kaggle.com

**Bloomberg, Structured Products Waterfall** June 2017-September 2018

*Senior Software Engineer*

- Designed and implemented a Spark-based infrastructure for high bandwidth data processing jobs
- Modeled time series data for 2.5 million asset-backed securities (ABS)
- Designed and implemented anomaly detection for ABS data passing through ETL and analytics pipelines

**Goldman Sachs, Operations Automation and Analytics Technology** July 2016-June 2017

*Technology Analyst*

- Helped build applications to consume, parse, & standardize market messages for fixed income instruments
- Designed & built a machine learning platform to create metrics & predictive models for Operations division

**Princeton University, Senior Thesis Research** September 2015-May 2016

*Senior Thesis Research*

- Investigated SVMs, Naïve Bayes, and ensemble methods for binary sentiment analysis on movie reviews
- Created a manually labeled corpus from 2004 Cornell IMDB data for subjective and summary sentences

**Goldman Sachs, Operations Analytics Strategies** June 2015-August 2015

*Summer Analyst*

- Led team working on automated invoice recognition using Tesseract and Python to process and automate template matching with noisy and tilted images containing structured and semi-structured text

**Princeton Laboratory for Energy Systems Analysis / CASTLE Lab** June 2014-August 2014

*Summer Research Intern*

- Created simulator in Java for the unit allocation problem of introduction of wind energy to the power grid
- Analyzed the role and performance of different classes of learning rates in reinforcement learning