

# Yilei Cai

☎ (669)248-6036 | ✉ [caiyilei3@gmail.com](mailto:caiyilei3@gmail.com) | [in yilei-cai](https://www.linkedin.com/in/yilei-cai)

## Skills

**Big Data** Spark, Flink, Kafka, Hive, Airflow, Oozie, Hadoop, Pig

**Programming** Java, Python, Scala, Ruby, C++, SQL, Shell, JavaScript, C, MATLAB, SML

## Education

### Carnegie Mellon University

M.S. IN COMPUTATIONAL BIOLOGY, SCHOOL OF COMPUTER SCIENCE

- Departmental Merit Fellowship (major GPA: 3.9/4.0)

*Pittsburgh, PA*

*Sep. 2013 - May 2015*

### Duke University

M.S. IN NEUROBIOLOGY, SCHOOL OF MEDICINE

- Withdrawal from PhD Program

*Durham, NC*

*Sep. 2011 - May 2013*

### Hong Kong University of Science and Technology

B.S. IN MOLECULAR BIOMEDICAL SCIENCES WITH MINOR IN MATHEMATICS, CHEMISTRY AND BIOLOGY

- University Scholarship, Lee Hysan Foundation Exchange Scholarship, Dean's List, First Class Honors

*Hong Kong*

*Sep. 2007 - Jun. 2011*

## Work Experience

### AirBnB

SOFTWARE ENGINEER, GROWTH AND TRAFFIC TEAM

- Developing offline and real time components for customer intelligence platform (Flink, Spark, Hive, Airflow)
- Rewrote the visitor-user mapping pipeline (Spark, Scala)
- Worked on data pipelines including cost ingestion (Ruby on Rails) and bot detection (Airflow, Hive)

*San Francisco, CA*

*Dec. 2018 - present*

### Yahoo!

SOFTWARE ENGINEER II, BIG DATA MACHINE LEARNING TEAM

- Implemented Spark based method to launch parameter server instances (Spark, Scala)
- Used TensorFlowOnSpark to generate word embeddings (word2vec), and optimize performance (Python)
- Made enhancements to parameter server based LDA (Scala)

*Sunnyvale, CA*

*Feb. 2018 - Nov. 2018*

SOFTWARE ENGINEER, GEMINI ADS TEAM

- Built offline model training and feature engineering pipelines (Pig, Oozie, HDFS, Python, Java, Shell, Hive)
- Added features for online ads scoring service (C++)
- Developed an internal reporting tool for ads data visualization based on the open source web application, Superset (Python, SQL, JavaScript)

*Nov. 2016 - Feb. 2018*

### Third Wave Systems

APPLICATIONS DEVELOPER

- Designed and implemented a new GUI de novo for a machining process simulating software
- Used C++ Open Cascade library for 3D geometry rendering and integrated it into GUI (java) using Java Native Interface (JNI)
- Added semi-automated report generation function (Java and C++)

*Minneapolis, MN*

*Jul. 2015 - Nov. 2016*

## Academic Projects

### Carnegie Mellon University

- Classified newspaper articles using machine learning models (Python, MATLAB)
- Designed and implemented a distributed Map/Reduce framework (Java)
- Designed and implemented a simple search engine (Python)
- Designed and implemented an urban transit system simulator (Java)
- Designed and implemented the game Scrabble with GUI (Java)
- Designed and implemented a social network analysis framework and its plugins (Java)

*Sep. 2013 - May 2015*

## Research Experience

### Duke University and HKUST

- Performed two-photon/confocal microscopy and biochemical experiments to study molecular neuronal plasticity

*Sep. 2009 - Dec. 2012*