

Yilei Cai

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

Objective and Skills

To obtain a position utilizing my skills in computer science:

PROGRAMMING LANGUAGES	Go, Java, Python, Scala, SQL, Ruby, C++, Shell, JavaScript, C, MATLAB, SML
TOOLS	Spark, Flink, Kafka, Hive, Airflow, Oozie, Hadoop, Pig
KNOWLEDGE	Big data processing (batch and stream), Machine Learning, Web service

Education

Carnegie Mellon University

Pittsburgh, PA

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

Sep. 2013 - May 2015

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

Duke University

Durham, NC

M.S. IN NEUROBIOLOGY, SCHOOL OF MEDICINE

Sep. 2011 - May 2013

- Withdrawal from PhD Program

Hong Kong University of Science and Technology

Hong Kong

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

Sep. 2007 - Jun. 2011

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

Work Experience

ByteDance

San Jose, CA

SOFTWARE ENGINEER, SECURITY TEAM

Dec. 2023 - present

- Improving data security platform's storage and traffic module (Go, Data Streaming, Hive)

TikTok

Mountain View, CA

SOFTWARE ENGINEER, DATA PLATFORM TEAM

Jun. 2020 - Sep. 2021

- Maintained TikTok's logging and data ingestion system (Flink, Kafka, Go, Python)
- Added features to the A/B testing platform (Python)

AirBnB

San Francisco, CA

SOFTWARE ENGINEER, GROWTH AND TRAFFIC TEAM

Dec. 2018 - Jun. 2020

- Built offline and real time components for customer data platform (Flink, Kafka, Spark, Scala, Hive, Airflow)
- Rewrote the visitor-user mapping pipeline, that infers a visitor's user id across different platforms (Spark, Scala)
- Maintained and improved data pipelines that prepare marketing summary and visits data. (Ruby on Rails, MySQL, Hive, Airflow)

Yahoo!

Sunnyvale, CA

MACHINE LEARNING ENGINEER, BIG DATA MACHINE LEARNING TEAM

Feb. 2018 - Nov. 2018

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

SOFTWARE ENGINEER, GEMINI ADS TEAM

Nov. 2016 - Feb. 2018

- Built offline model training and feature engineering pipelines for ads ranking (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)

Third Wave Systems

Minneapolis, MN

APPLICATIONS DEVELOPER

Jul. 2015 - Nov. 2016

- 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 a semi-automated report generation feature (Java and C++)

Academic Projects

Carnegie Mellon University

Sep. 2013 - May 2015

- Designed and implemented a distributed Map/Reduce framework (Java)
- Designed and implemented a simple search engine [link] (Python)

Research Experience

Duke University and HKUST

Sep. 2009 - Dec. 2012

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