

Yilei Cai

☎ (669)248-6036 | ✉ caiyilei3@gmail.com | 🌐 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

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

ByteDance

SOFTWARE ENGINEER, SECURITY TEAM

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

San Jose, CA

Dec. 2023 - present

TikTok

SOFTWARE ENGINEER, DATA PLATFORM TEAM

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

Mountain View, CA

Jun. 2020 - Sep. 2021

AirBnB

SOFTWARE ENGINEER, GROWTH AND TRAFFIC TEAM

- 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)

San Francisco, CA

Dec. 2018 - Jun. 2020

Yahoo!

MACHINE LEARNING ENGINEER, BIG DATA MACHINE LEARNING TEAM

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

Sunnyvale, CA

Feb. 2018 - Nov. 2018

SOFTWARE ENGINEER, GEMINI ADS TEAM

- 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)

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

Minneapolis, MN

Jul. 2015 - Nov. 2016

Academic Projects

Carnegie Mellon University

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

Sep. 2013 - May 2015

Research Experience

Duke University and HKUST

Sep. 2009 - Dec. 2012

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