

Kehang Han

(617)694-7029
11 Peabody Terrace
Cambridge, MA 02138

hkh12@mit.edu
www.kehanghan.com
www.github.com/kehang

Current Position **Staples, Inc.**, Framingham, MA
Data Scientist

July 2018 - Present

- Deep learning based Image Product Detection service: FasterRCNN model, Active learning, Azure Cloud
- Product Matching service: ETL pipeline, ordinal classification, Azure Cloud
- Intelligent Inventory Management system: IoT hardware, UI design, Azure Cloud, API hub
- Open source software - model2service: conveniently converts ML models into API services
- Other in-house services: Named Entity Recognition, General Sequence Classification Platform, Warehouse Simulator, Robotics Optimization.

EDUCATION **Massachusetts Institute of Technology**, Cambridge, MA
PhD
Advisor: William Green

June 2014 - May 2018

- Chemical Engineering (major) and Computer Science (minor)
- Overall GPA: 4.9/5.0
- Doctoral Dissertation: Enabling Automatic Generation of Accurate Kinetic Models for Complicated Chemical Systems

Tsinghua University, Beijing, China
Bachelor of Science

Sep 2008 - Jul 2012

- Chemical and Biological Engineering, Rank: 1/120

MACHINE LEARNING **Self-Evolving Molecular Property Estimator**
Lead Developer

May 2016 - May 2017

- Designed and implemented Molecular Convolution Neural Networks (MCNN) that effectively featurize molecule graphs
- Created open source software - DataDrivenEstimator: utilize MCNN to predict molecular properties such as enthalpy
- Constructed data pipeline enabling self-evolution of the DataDrivenEstimator models

Amazon Product Review Summarizer (SRS, srs.mit.edu)
Lead Developer

Mar 2016 - Sep 2016

- Led a 3-member team at MIT
- Built an NLP application (srs.mit.edu) in Python, Flask and MongoDB
- Constructed four-step pipeline: review scraping, aspect classification, sentiment scoring and front-end interaction
- Developed product-aspect classifiers using MaxEntropy and Word2Vec

DATA SCIENCE **Supply Chain Planner in Shell Oil**, Hamburg, Germany
Data Science Intern Jun 2015 - Sep 2015
Supervisor: Marc Thomas

- Streamlined price data processing for supply chain optimization tool
- Created analysis and visualization tools using R for extreme scenario study

SOFTWARE DEVELOPMENT **Reaction Mechanism Generator (RMG, rmg.mit.edu)**
Lead Developer Jan 2013 - Present

- Led a team with 10 people at MIT and Northeastern University
- Applied high performance computing (memory reduction, parallelism) to RMG
- Maintained cross-platform (Linux, OSX, Windows) compatibility
- Constructed Continuous Integration Test Platform (RMG-tests) for RMG

PROtein Design Algorithm (PRODA)
Developer Sep 2011 - Jul 2012

- Development team member of PRODA, a software for designing protein using C
- Designed and implemented a heuristic optimization algorithm for MILP
- Achieved 20 times speed-up for large enzyme systems

**AWARDS
HONORS**

Energy Analytics

- MIT Energy Initiative (MITEI) Fellowship in 2016
- 3rd Prize (3/30 teams) of MIT Energy Hackathon in 2015
- 3rd/300+ in “Mitsui Chemicals Cup” Process Design Contest in 2011

Mathematics

- 2nd Prize in American Mathematical Contest in Modeling in 2011
- 1st Prize in Chinese National Mathematical Contest in Modeling in 2010

Physics & Chemistry

- 1st Prize in Beijing Physics Contest in 2009
- 1st Prize in Chinese National Olympic Chemistry Competition in 2007

**SELECTED
PUBLICATIONS**

- **K. Han**, W.H. Green, and R.H. West, *On-the-fly pruning for rate-based reaction mechanism generation*, Computers & Chemical Engineering
- YP Li, **K. Han**, C.A. Grambow, W.H. Green *Self-Evolving Machine: A Continuously Improving Model for Molecular Thermochemistry*, The Journal of Physical Chemistry A
- X. Huang, **K. Han**, and Y. Zhu, *Systematic optimization model and algorithm for binding sequence selection in computational enzyme design*, Protein Science
- **K. Han**, W.H. Green, *A Fragment-Based Mechanistic Kinetic Modeling Framework for Complex Systems*, Industrial Engineering Chemistry Research

SKILLS

Languages: Python, C/C++, Java, R, Matlab, PHP
Database: SQL, mongoDB, PostgreSQL
Data Science: pandas, scikit-learn, matplotlib, Spark
Deployment: Azure, AWS, server admin