Myungsun Kang

241 River street, Cambridge, MA 02139
sunny.myungsun.kang@gmail.com
https://www.linkedin.com/in/myungsun-sunny-kang/

Senior Data scientist, Ph.D. in computational vaccine design

KEY QUALIFICATIONS

Senior Data scientist with experience productionalizing machine learning pipeline and developing new machine learning models at Forbes 500 e-commerce (Wayfair)

Research scientist at Covid-19 tech consortium (PathCheck), developing neural networks for proximity sensing via Bluetooth signal

6 years of domain expertise in HIV vaccine science, modelling and statistical analysis at MIT, which led to publications of **2 joint first author articles in top peer-reviewed journals**

EXPERIENCE

2019 Mar- **Senior data scientist**, Wayfair, Algorithms and Analytics

- Evaluated the impact of Wayfair's new logistics program for large parcel delivery and leveraged the result to create a model that predicts the expected impact for a new product joining the program
- Productionalized an upgrade to a major machine learning pipeline that predicts Wayfair's profitability for million orders per day by leveraging HIVE, PySpark and Airflow
- Built and productionalized a machine learning models that predict the margin from a special category of Wayfair's product, which increased accuracy by 65 percent in comparison to the previous model

2020 May- **Pro Bono Research scientist**, PathCheck Foundation

- Built 1D Convolutional neural networks that predict the distance between two devices from the time series records of bluetooth and various sensors
- Engineered uniform length feature set from time series stamps of Bluetooth signals and various sensors published by NIST
- Built three strategies to tackle the overfitting and grid searched for hyper parameters such as epoch size, learning rate schedule and batch size
- Took the third place in the competition held by NIST and submitted a paper to arXiv

2018 Sep-Nov **Fellow**, Insight Data Science

- Built a web app that recommends anti-depressants pills tailored to patients' symptoms
- Extracted side effects from patient survey data using topic modeling, and infered their prevalence from the corpus of patient experience
- Engineered features from 17,000 subreddit comments collected through Pushshift Reddit API using NLP methods including TF-IDF, word2vec and sentiment analysis
- Built a recommendation system that is trained on words associated with positive experiences for a given drug. Employed Logistic regression and linear SVC for the classification
- Built an inter-active user interface with Flask, Bootstrap and AWS

2012 -2019 Feb **Graduate researcher**, Chemical Engineering, MIT

- Built modeling and analytical tools to predict immune response from HIV vaccine prototypes
- Employed non-linear regression to built a time-dependent deterministic nonlinear differential equation model in Matlab, which can predict serum Ab production upon vaccination
- Built a stochastic model of Ab response by implementing Tau-leap gillespie algorithm with partial deterministic approximation, which led to 50X speed enhancement with <1 percent accuracy tradeoff
- Simulated HIV vaccine prototypes and proposed prospective candidates, which is now being tested in non-human primates

EDUCATION

2012 –2019 **Massachusetts Institute of Technology (MIT)**
Doctor of Philosophy candidate, Chemical Engineering, Institute for Medical Engineering & Science
(Minor: Statistics and Computer Science)
Master of Science in Chemical Engineering Practice, Chemical Engineering

2008–2012 **Korea Advanced Institute of Science and Technology (KAIST)**
Bachelor of Science, Chemical Engineering, Minor: Biology, *summa cum laude*

PUBLICATIONS

Google Scholar profile: https://scholar.google.com/citations?user=0R3DecUAAAAJ&hl=en

Sheshank Shankar*, Ayush Chopra*,Rishank Kanaparti*, **Myungsun Kang***, Abhishek Singh, Ramesh Raskar, "Proximity Sensing for Contact Tracing", arXiv, 2020 Aug **submitted**

Hok Hei Tam*, Mariane B. Melo*, **Myungsun Kang***, Jeisa M. Pelet, Vera M. Ruda, Maria H. Foley, Joyce K. Hu, Sudha Kumari, Jordan Crampton, Alexis D. Baldeon, Rogier W. Sanders, John P. Moore, Shane Crotty, Robert S. Langer, Daniel G. Anderson, Arup K. Chakraborty, Darrell J. Irvine, "Sustained antigen availability during Germinal Center initiation enhances antibody responses to vaccination", Proceedings of the National Academy of Sciences, 201606050 (*equal contributors) **101 accumulated citations to date**

Kang M, Eisen TJ, Eisen EA, Chakraborty AK, Eisen HN (2015), "Affinity inequality among serum antibodies that originate in lymphoid Germinal Centers", PLoS ONE 10 (10): e0139222. doi:10.1371/journal.pone.0139222 **11 accumulated citations to date**

SKILLS

Technical expertise	Immunology, Vaccine design, Discrete stochastic simulation, Machine learning, Causal inference, NLP, Deep Learning, Exploratory statistics, Inferential statistics, Bayesian statistics
Programming Languages	Python, PySpark, Pytorch, Tensorflow, Keras, SciPy, NumPy, Pandas, Seaborn, StochPy, C, Bash
Query Languages	SQL, HIVE
Applications	Git (source control), Docker, Airflow, BigQuery, GCP
Languages	English and Korean (fluent)