Patrick Lam

pkphlam@gmail.com
www.patricklam.org
www.linkedin.com/in/pkphlam

EXPERIENCE

Meta (formerly Facebook)

Senior Research Scientist (Core Data Science team) 2019 - Present

- **Data For Good Disaster Maps**: Collaborated with the Meta Data For Good team to develop and share real-time, privacy-preserving geographical datasets from Facebook location and activity data, aimed at improving the response of disaster relief organizations. This involved developing statistical methodology to create the datasets, engineering data pipelines using SQL, Python, and Dataswarm (an internal version of Apache Airflow) to deliver daily data to partners, and maintaining the data pipelines as part of an engineering on-call rotation.
  - Built new datasets to measure economic business recovery in disaster areas. The resulting datasets and insights were used by external partners to respond to crises such as COVID-19, the War in Ukraine, and the 2023 Turkey-Syria earthquake.
  - Optimized multiple existing data pipelines in the Meta Data for Good portfolio, resulting in compute cost savings of over 90% and significantly reduced data downtime (greater than 95% on-time delivery).
- **High Resolution Settlement Layer (HRSL)**: Built machine learning classifiers to predict physical building type (residential vs. non-residential) based on satellite and Facebook data to increase accuracy of fine-grained human population estimates as part of the Facebook HRSL mapping project.
- **Social Capital Lab**: Established engineering standards and created data pipelines to measure social capital and economic mobility in the UK as part of the newly launched Social Capital Lab initiative.
- **Well-Being Products on Facebook**: Analyzed the impact of Facebook feature launches on emotional well-being outcomes using advanced causal inference methods (principal stratification) and machine learning models.

Thresher (acquired by Two Six Technologies in 2022)

Lead Data Scientist and Co-Founder 2015 - 2019

- Designed and implemented the Thresher QuickCode algorithm that utilized machine learning to suggest labeling rules for training data in text classification.
- Served as Principal Investigator for several DARPA-funded projects on big data, social media manipulation, and the spread of information.
- Programmed in Python for various functions such as web data scraping, API data downloading, text data processing using NLP techniques, Elasticsearch data indexing, querying, and aggregation, and applied both supervised and unsupervised machine learning models.
- Collaborated with analysts to create data visualization tools and gain insights from data.

Minerva Schools at KGI (now Minerva University)

Assistant Professor of Computational Science (part-time) 2018 - 2019

- Led courses on "Causality and Common Sense Reasoning in Machine Learning" and "Computational Data Analysis and Modeling in Social Sciences."

Australian Conservation Foundation

Statistical Consultant (freelance) 2018 - 2019

- Conducted field experiments in Australia to evaluate the effectiveness of campaign canvassing techniques in increasing engagement on environmental issues.

International Finance Corporation

Statistical Consultant (freelance) 2012 - 2017

- Evaluated the impact of IFC programs on development outcomes for the IFC SME Finance Initiative and other projects.

Australian Council of Trade Unions

Statistical Consultant (freelance) 2013 - 2014

- Conducted data analysis and field experiments to evaluate the effectiveness of campaign canvassing techniques on vote outcomes for the 2013 Australian Federal Election.

PUBLICATIONS

- King, Gary, Patrick Lam, and Margaret E. Roberts. 2017. "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." *American Journal of Political Science* 61 (4): 971-988

EDUCATION

Harvard University

Ph.D. in Political Science and A.M. in Statistics 2013

- Dissertation: Estimating Individual Causal Effects
- Committee: Gary King (chair), James Alt, Adam Glynn, Arthur Spirling

University of California, Los Angeles

B.A. in Political Science 2006

INVITED TALKS

- "QuickCode: Label your Training Data Fast and Transparently." *PAPIs*, Boston, October 17, 2018
- *15th National Security Seminar*, Singapore, October 17, 2017
- *Computational Simulation of Online Social Behavior (SocialSim)*, Defense Advanced Research Projects Agency, October 11, 2017 and July 17, 2018.
- "National Security & Intelligence Applications of Text Analytics." *Topics in Cyber Security and the Internet*, University of Central Florida, March 26, 2015.
- "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." *31st Annual Meeting of the Society for Political Methodology*, University of Georgia, July 24, 2014.

TEACHING

Minerva Schools at KGI

Assistant Professor of Computational Science   2018 - 2019

- Designed and instructed an undergraduate tutorial on Causality and Common Sense Reasoning in Machine Learning, exploring statistical and AI approaches to reasoning and causality.
- Taught undergraduate tutorials on Computational Data Analysis and Modeling in Social Sciences, guiding students in the use of statistical techniques and computational tools to analyze social data.

Harvard University

Instructor and Teaching Fellow   2008 - 2011

- Designed and instructed an informal graduate course covering topics in Bayesian statistics (conjugacy, hierarchical models, Markov Chain Monte Carlo methods, convergence, and model checking) using R. Previously served as a Teaching Fellow for a similar graduate course.
- Served as a Teaching Fellow and Instructor for a graduate math refresher course for incoming political science graduate students, covering calculus, probability, and R.
- Led graduate-level courses as a Teaching Fellow on advanced political methodology topics, including generalized linear models, causal inference, matching methods, and missing data imputation using R.
- Served as a Teaching Fellow for an undergraduate introductory course on basic statistical analysis using SPSS.

Massachusetts Institute of Technology

Teaching Assistant   2011

- Served as a Teaching Assistant for a graduate course on linear regression and causal inference.

SKILLS

- **Languages/Tools**: Python, R, SQL, Dataswarm (internal Facebook version of Apache Airflow), Elasticsearch, Git
- **Statistical Expertise**: causal inference in observational data, Bayesian statistics, supervised and unsupervised machine learning, text analysis, missing data methods