

Website: padtai.ca
LinkedIn: [linkedin.com/in/taicai](https://www.linkedin.com/in/taicai)
GitHub: github.com/padtai

— TIANYI (TAI) CAI —
Brookline MA
Canadian Citizen with H-1B

Updated 01/2019
tcai@mail.harvard.edu
(857) 998-8619

SUMMARY

PhD training	Bayesian statistics (multilevel models, causal inference, model averaging, nonparametric theory) Classical statistics (hypothesis testing, mixed models, regularization, asymptotic theory)
Current interests	Artificial neural networks, statistical properties of machine learning methods, Bayesian methods for machine learning
Comfortable using	Python (NumPy, pandas, matplotlib, scikit-learn, XGBoost, Flask, Beautiful Soup) R, SQL, Jupyter Notebook, Amazon S3, Amazon EC2, \LaTeX , Bash, git, tmux
Some exposure	Tensorflow, Keras, Spark, Java, C++, HTML, CSS, Javascript

EXPERIENCE

Data Scientist , BitSight	<i>Boston MA</i>	10/2017–Present
<ul style="list-style-type: none">Oversee all data science requirements for third party risk management projects by collaborating with cross-functional teamsPartner with product, sales, and marketing managers to devise data-driven strategies from unorthodox data structuresDesign studies to evaluate the association and causality of relationships pertaining to efficacy of products, impact of extraneous events, and influence of internal interventionsProgram Python scripts and modules for data storage, collection, cleaning, analysis, and visualizationLead reading groups covering advanced topics in forecasting methods, prediction intervals, and model evaluation metricsSupport data science team members in ad hoc statistical tasks		
Data Science Fellow , Insight Data Science	<i>Boston MA</i>	05/2017–09/2017
<ul style="list-style-type: none">Generated idea for predicting supply/demand at Hubway bike-sharing stations in Boston, despite a lack of labeled dataConsolidated and cleaned multiple data sources to tally labeled information for 200 bike stations over 11 million time pointsPredicted real-time supply/demand for each bike station using various machine learning and statistical modelsCreated website to visualize directions and to predict bike availability for user input origins and destinationsProvided mentorship and feedback for subsequent cohorts of fellows		
PhD Student, Research Assistant , Harvard University	<i>Cambridge MA</i>	08/2012–09/2017
<ul style="list-style-type: none">Conducted original statistical research on applied problems in mental disorders, health care policy, and end-of-life careDeveloped R code for handling datasets with complexities such as sampling bias, misclassified outcomes, correlated outcomes, hierarchical structures, and confoundingTaught graduate-level labs with topics ranging from introductory statistics to seminar topics in Bayesian nonparametrics, decision theory, and sequential methodsSelected to tutor fellow PhD students for the biostatistics written qualifying exam		
Full-time Co-op Work Semesters , University of Waterloo	<i>Waterloo ON</i>	05/2008–12/2011
Actuarial Analyst, Biometric Research, Enterprise Risk Management Munich Re	<i>Toronto ON</i>	01/2011–12/2011
Actuarial Analyst, Group Benefits Pricing, Segregated Fund Valuation, Manulife	<i>Waterloo ON</i>	09/2009–08/2010
Pension Administrator, Hewitt Associates	<i>Toronto ON</i>	01/2009–04/2009
Database Analyst, Logitech	<i>Mississauga ON</i>	05/2008–08/2008

EDUCATION

PhD, Biostatistics , Harvard University	<i>Cambridge MA</i>	2017
<ul style="list-style-type: none">Thesis: Statistical Methods for the Analysis of Observational Data with Multiple Correlated OutcomesAdvisors: Tianxi Cai, Francesca Dominici		
Bachelor of Mathematics , University of Waterloo	<i>Waterloo ON</i>	2012
<ul style="list-style-type: none">Honours Actuarial Science/Finance Option, Honours Statistics, Co-operative Program		
Associate of the Society of Actuaries , Society of Actuaries	<i>Schaumburg IL</i>	2012