John Cai

jjcai@alumni.princeton.edu | linkedin.com/in/caijohn | github.com/johncai117 | website: johnc.ai

Experience

- **Snap Inc.** — Santa Monica, CA
  Applied Research Scientist, Engineering — *Nov 2021 - Present*
  - Led a team of 4 applied scientists to characterize the effect of app performance on engagement using causal machine learning and meta-learners. Built a SQL query builder API to operationalize a new app performance metric.
  - Quantified ads lift using machine learning and built models to rank house ads based on predicted user retention. Results were used to allocate over $XXm worth of annual house ads budget.
  - Performed user modelling using generalized additive models to predict and drive long-term user retention.
  - Developed new algorithms for causal graph learning using shrinkage estimation to map out how different metrics relate to each other. Insights were used by VPs and CEO for decision making and prioritization
  - Devised a new method to detect, surface and characterize heterogenous treatment effects of recommender systems.
  - Built a self-serve tool in Shiny to diagnose variance overestimation in new randomization IDs for A/B testing
  - Launched an anomaly detection tool in Google Cloud to detect abnormally high cloud usage using machine learning and time series analysis. By reducing cost anomalies, the model is able to save up to $XXm a year.
- **Singapore Armed Forces** — Singapore
  Corporal — *Aug 2020 - Oct 2021*
  - Completed mandatory military service. Created Power BI dashboards to monitor and analyze manpower trends.
- **Government Technology Agency** — Singapore
  Machine Learning Intern, Artificial Intelligence Platforms — *May 2020 - Jul 2020*
  - Deployed a machine learning recommender system that would improve click through rate by around 3%.
  - Created a forecasting model in PySpark and Databricks to predict the COVID labor market.
- **Smart Nation and Digital Government Office** — Singapore
  Digital Strategy Intern — *Jun 2018 – Aug 2018*
  - Analyzed financial transaction data to identify policies that can reduce infrastructure costs by $10m/year

Education

- **Princeton University** — Princeton, NJ
  Masters, Financial Mathematics; GPA: 4.00/ 4.00 — *Aug 2019 – Jun 2020*
  - **Relevant Courses**: Machine Learning and Pattern Recognition, Natural Language Processing, Deep Learning Networks (PhD), Advanced Computer Vision (PhD), Reinforcement Learning (PhD), Statistical Analysis of Financial Data, Computational Finance in C++, Stochastic Calculus, Financial Econometrics
- **University of Cambridge** — Cambridge, UK
  Bachelors, Economics; Triple First Class Honours — *Oct 2016 – Jun 2019*
  - **Relevant Courses**: Econometrics, Mathematics, Statistics, Functional Programming (PKU Exchange)
  - **Dissertation**: Predicting Recessions using Uncertainty-Aware Natural Language Processing and Factor Models

Research Publications and Projects

- **Domain Agnostic Meta Score-based Learning (CVPR 2021 L2ID)**: Introduced a transfer-learning algorithm using a graph neural network in PyTorch. Beats state-of-the-art methods on medical and satellite images.
- **Cross-Domain Few-Shot Learning (CVPR 2020 VL3)**: 2nd in the CDFSL Challenge against established AI teams from Alibaba and Didi. Developed a meta learning algorithm that generalized well on diverse image domains.
- **NLP Event-Driven Stock Price Prediction**: UUtilized OpenIE to parse relation tuples followed by a deep CNN to extract short-term and long-term features from news articles, which are then used to predict stock price movements.

Skills and Interests

- **Programming:** Python, R, C++, SQL, Javascript/ Node.js, Haskell, MATLAB, Stata, Bash
- **Tools:** PyTorch, TensorFlow, Keras, PySpark, OpenCV, NLTK, FastAI, HuggingFace, Heroku, Docker, Kubernetes, sklearn, NumPy, Pandas, Shiny, Arrow, seaborn, ggplot2, Flask, React, Gatsby, PostgreSQL, VSCode, Jupyter, Git
- **Cloud Computing:** AWS (EC2, S3, Lambda), GCP (BigQuery, Dataproc, Compute Engine), Databricks
- **Academic Interests**: Graph Learning, Computer Vision, Applied Statistics, Causal Inference, MLOps
- **Other Interests**: Hiking, Climbing, Cycling, Boxing (Trained with University team), Theatre