Jay Zern Ng
https://jayzern.github.io
155 Claremont Avenue, New York, NY 10027
jn2717@columbia.edu ⋄ +16463014870

Education

- **Columbia University** — New York, NY
  M.S. in Data Science (Software Systems & Machine Learning); GPA: 3.67/4.00 — *Aug 2019 – Dec 2020*
  - Operating Systems, Programming Languages & Translators, Computer Networks, Intro to Databases, Applied Deep Learning, Applied Machine Learning, Exploratory Data Analysis & Visualization, Statistical Inference & Modeling, Probability & Statistics, Machine Learning with Probabilistic Programming
- **University of Warwick** — Coventry, UK
  B.S. in Data Science (Bayesian Statistics); First Class Honors: 80.2% — *Oct 2016 – Jul 2019*
  - Bayesian Forecasting, Bayesian Statistics, Stochastic Procceses, Multivariate Statistics, Neural Networks & Reinforcement Learning, Machine Learning, Mathematical Analysis I, II & III, Algorithms & Data Structures, Mathematical Statistics A & B

Awards: Outstanding Dissertation Award (Warwick), Project Contest winner for Intro to Databases (Columbia), 2nd place Kaggle competition for Machine Learning (Warwick), Top 7 Data Science Hackathon (Columbia), 3rd place Barclays Hackathon

Programming

Languages › C, C++, Java, Python, SQL, R, Typescript, Javascript, OCaml
Software › AWS, GCP, Docker, Kubernetes, Jenkins, Advanced UNIX, Angular, React.js, Node.js, LLVM
Data › Scikit-learn, TensorFlow, PyTorch, D3.js, Tidyverse (ggplot2, dplyr), Hadoop, Spark

Experience

- **Boost** — Kuala Lumpur, MY
  Software Engineer Intern — *May 2020 – Sep 2020*
  - Boost is a lifestyle e-wallet app with 10 million+ users. Built internal registration systems and microservices using Angular, RxJS and Node.js. Deployed and monitored CI/CD pipelines using Docker, Kubernetes, Jenkins, AWS.
- **Columbia Law School** — New York, NY
  Data Engineer (Research Assistant) — *Sep 2019 – Apr 2020*
  - Wrote a data engineering ETL pipeline for 100GB+ of JSON/XML text data into a PostgreSQL database. Developed a tool to detect academic citations using Regular Expressions and spaCy Named Entity Recognition.
- **Barclays** — London, UK
  Cybersecurity Intern — *Jun 2018 – Aug 2018*
  - Monitored static application security dashboards using Checkmarx and restructured the team's database schema using SQL. Developed a prototype chatbot that recommends Barclaycard products using Django and DialogFlow as a side project.
- **VLT Labs** — Kuala Lumpur, MY
  Software Engineer Intern — *Jul 2017 – Sep 2017*
  - VLT Labs is a venture builder based in Southeast Asia. Built chatbots using Angular, Node.js, MongoDB and spaCy to automate customer support. Implemented a frontend framework for an online retailer using object-oriented CSS.

Selected Projects

- **TI-Lisp** Type Inference Lisp is a new functional programming language written in OCaml, C++ and LLVM, inspired by the Scheme language from MIT. It combines the power of a Lisp language and type checking; providing the best of dynamic and static typing languages. Implemented features such as macros, in-built functions, code optimization and garbage collection.
- **Bloomberg: Automatic event understanding using Deep Learning** Columbia Data Science Capstone project. Understanding events using NLP based semantic role labelling and thematic fit; on the Rollenwechsel-English (RW-eng) corpus. Implemented attention based models, non-sequential (wide-and-deep, residual networks) and sequential (LSTMs, RNNs) architectures.
- **Kitchen OS** For my Operating Systems class, I completed a series of kitchen-themed projects based on the Linux source tree in C and UNIX. Implemented a "Freezer" round-robin scheduler system for multi-processor CPUs, a "Fridge" kernel key-value store with blocking/non-blocking syscalls, and a "Pantry" filesystem with fully featured read/write/delete operations.
- **Bayesian Online Changepoint Detection** My bachelor's thesis on time series analysis. Implemented the Log Gaussian Cox Process model in Python, with applications on weather forecasting data. Extended the model using sparse variational inference for faster inference and multitask learning for modeling covariances.