Helen Ngo

mathemakitten.dev
416 806 8161 • helen.ngo14@gmail.com

experience

Cohere

Research Engineer, Safety Team Lead — January 2020 - March 2022

- Founding engineer; designed and implemented core model training framework in Tensorflow used to train and evaluate multi-billion parameter Transformer language models; later rewritten in JAX
- Designed and implemented safety system using language models for dataset curation & filtration of toxic and undesirable web data resulting in safer models which minimize alignment taxes
- Led safety research team responsible for building technical measurement and mitigation systems
- Wrote public-facing documentation communicating best practices around responsible use and collaborated with product managers to write best practices around product safety and policy

AI Index, Stanford Human-Centered Artificial Intelligence (HAI)

Affiliated Researcher — November 2021 - March 2022

- Conducted meta-analysis on research literature surveying technical metrics on bias and fairness, ethics, and conference publication trends for the 2022 AI Index Report
- Authored featured chapter (technical AI ethics)

Recurse Center

Fellowship Recipient — Fall 2019

- Researched misinformation and the integrity of our information ecosystem on Twitter. Built dashboards on Canadian elections & disinformation in Hong Kong, incorporating natural language processing tools

Dessa (acquired by Square Inc.)

Machine Learning Engineer — January - December 2019

- Designed, built, and shipped production machine learning systems for Fortune 100 enterprises at multi-terabyte scale across a variety of industries using Python, Tensorflow, Spark, Docker, and SQL
- Key projects: credit card fraud detection for top-3 U.S. creditor, social media diversity & inclusion brand sentiment analysis tool for Diversio, consulted on credit card fraud system for Canadian bank

Sidewalk Labs & Waterfront Toronto

Research Fellow — March - December 2018

- Selected as 1 of 12 Fellows as part of Sidewalk Toronto's mission to build a next-generation smart city
- Consulting on technology and policy with the goal of improving urban life for all. Report: link

Bell

Data Scientist, AI Labs team — June 2016 - December 2018

- Built, trained and deployed deep learning classifier or call centre prediction; 20% lift over baseline
- Regression & tree-based modeling for churn, pricing sensitivity, iconic device launches, ad targeting
- Integrated models into production environment creating 100M+ scores weekly

Helen Ngo

mathemakitten.dev
416 806 8161 • helen.ngo14@gmail.com

research

No News is Good News: A Critique of the One Billion Word Benchmark

Helen Ngo, João GM Araújo, Jeffrey Hui, Nicholas Frosst

Accepted to the Data-Centric Workshop @ NeurIPS 2021. Selected as a lightning talk.

Mitigating Harm in Language Models with Conditional-Likelihood Filtration

Helen Ngo, Cooper Raterink, João GM Araújo, Ivan Zhang, Carol Chen, Nicholas Frosst

Predicting Twitter Engagement with Deep Language Models

M Volkovs, Z Cheng, M Ravaut, H Yang, K Shen, JP Zhou, A Wong, S Zuberi, I Zhang, N Frosst, **H Ngo**, C Chen, B Venkitesh, S Gou, A.N. Gomez

Proceedings of the Recommender Systems Challenge 2020

awards, writing, miscellaneous

- Writer-in-Residence @ The University of Western Ontario (2015-2016), selected from 400 applicants to be first STEM student named to Writer-in-Residence post
- **Book**: Reimagining ChinaTOwn: Speculative Stories from Toronto's Chinatown (Mawenzi House 2023)
- **Community organizing:** Toronto Women's Data Group organizer (2017 - 2020)
- **Projects**: @whalefakes (finetuned GPT-2 Twitter bot), Wolfram Beta (Transformers for mathematics)
- **Awards:** OpenAI Scholar 2020 (declined), Recurse Center fellowship 2019, ACM Recsys Challenge 2019 (top 3% of teams worldwide), VentureBeat Women in AI Award finalist 2019, PyData NYC Scholarship 2019 (declined), Dockercon Scholarship 2018

education

The University of Western Ontario — 2016

Honors Specialization, Mathematics (B.Sc), graduated on Dean's Honor List
minor in Writing Studies

Awards: Alfred Poynt Poetry Award, Marie Smibert Writing Prize