Nathan Kim

(+1)-650-505-4069 | nathangk@stanford.edu | linkedin.com/in/nathangk | nathankim7.github.io

EDUCATION

Stanford University **Stanford, CA**
BS, Symbolic Systems; MS, Computer Science; Minor in Classical Languages *Expected June 2024*
Key Coursework: NLP w. Deep Learning (Best Project), Machine Learning (Best Project), Deep Generative Models

RESEARCH

Holistic Evaluation of Language Models **Transactions on Machine Learning Research (TMLR)**
Authored with Stanford CRFM
This paper introduces HELM, a **living benchmark** for foundation models designed to capture the full space of LLM use cases in a principled manner. **Individually**, I implement the International Corpus of English (ICE) as an evaluation on language modelling fairness across English dialects of varying prestige, and argue for its importance in an assessment of bias expressed by LLMs. I also discuss the results of our suite of linguistic knowledge evaluations more generally.

GLARE: Infilling Language Models for Textual Adversarial Attacks **Eval4NLP @ AACL–IJCNLP 2022**
Nathan Kim, Ryan Chi*, Patrick Liu, Zander Lack, Ethan Chi*
Textual adversarial examples can expose serious security vulnerabilities in NLP models by illustrating where small perturbations in input texts can drastically alter their predictions. We find that the GPT-2 adapted to perform non-causal infilling (Donahue et al. 2020) outperforms existing adversarial example generation methods in fluency, semantic preservation, and attack success rates.

WORK EXPERIENCE

Portalform (now Fiber) (YCombinator W23) **January 2023—March 2023**
Core Engineering Intern

- **Bootstrapped** a next-generation software layer for SaaS to automate 3rd-party client data ingestion.
- **Designed** a reader service to sync client databases with Amazon + Shopify. Built on Prisma, PostgreSQL & Express.

Canal **June 2022—September 2022**
Frontend Engineering Intern

- Contributed **18.5% of frontend code** written at Canal during internship period, including **redesigned invite interface**
- **Initiated/designed major refactor** of Proposal screens and provided **key input on data model** for V2 Proposal system

PROJECTS & VOLUNTEER WORK

Causal Alignment for Controlled Text Generation **October 2023—**
Lead Researcher, advised by Zhengxuan Wu, Christopher Potts

- Thesis research on developing prompt-free one-click controls for natural language generation (NLG) with LLMs
- **Expands** on theory of IIT (Geiger et al. 2022) to control for causal variables (sentiment, phrase structure) in generated sequences of infinite length
- Matched accuracy + fluency of existing CTG methods with **zero** new trained parameters
- **Initiated/designed** implementation of all methods, to appear in open-source ML library Pyvene

Candid **April 2023—June 2023**
Mobile Engineer

- Freelance development for a Stanford social media company intent on making digital mental health work for real.
- **Stack**: React Native + iOS Native extension, AWS Amplify, Firebase

AWARDS

- Gold Medal Division, 2020 Asia Pacific Linguistics Olympiad
- Silver Medalist, 2019 International Linguistics Olympiad
- Student Honor Roll, 2018 Canadian Computing Competition

SKILLS

- Pytorch, Pandas, R, HTML/CSS. React.js, Flask, Node.js, Java, React Native, C++, Rust