Brian Yu

For more information, please visit my personal website bri25yu.github.io by scanning the QR code.

ABOUT ME

I'm a highly motivated and dedicated **expert research engineer** with extensive industry experience. My technical expertise includes programming in Python, PyTorch applications, deep learning modeling, and Hugging Face Transformers. My theoretical expertise includes cognitive language modeling, vector calculus, and optimization. I excel at consuming literature and introducing clarity into systems and projects. Overall, I'm a happy and easygoing collaborator who rises to the challenge. Thanks for reading!😊

ACADEMIA

Deep Learning Language Modeling research
University of California, Berkeley. Berkeley, CA / Oct 2021 - Present
Principal investigator: Kurt Keutzer. Work done at Berkeley Artificial Intelligence Research (BAIR)

- Invented data efficiency methods for low resource translation with a focus on classical Tibetan
- Investigating language understanding model improvements through simple data transformations
- Improved mT5 Tib to Eng translation from 23.5 BLEU to 24.6 BLEU (publication in progress)

WORK EXPERIENCE

Deep Learning Research Engineering Internship - Alexa Question and Answering
Amazon.com, Inc. Seattle, WA (work done at Palo Alto location) / Jun - Aug 2022

- Created an efficient framework to rigorously ablate over large language models (LLMs)
- Introduced clarity in Alexa's QA evaluation research and opened up multiple research avenues
- Invented data efficient methods that reduced train set size by 16x and training time by 20x
- Improved long tail multilingual QA quality classification by 31.0% (Amazon internal, no publication)

Data Scientist / Research Engineering Internship - Alexa Question and Answering
Amazon.com, Inc. Seattle, WA (work done at Palo Alto location) / Jun - Aug 2021

- Created scalable, maintainable, and efficient codebases for performing various experiments
- Delivered results and iterated quickly through experimental configurations to impact end users
- Improved long tail QA quality classification F1 by 14.7% (Amazon internal, no publication)

EDUCATION

Master of Science in Electrical Engineering and Computer Science
University of California, Berkeley. Berkeley, CA
Aug 2022 - May 2023 (Expected graduation date)
Technical GPA: 4.00

Bachelor of Science in Electrical Engineering and Computer Science
University of California, Berkeley. Berkeley, CA
Aug 2018 - May 2022
Technical GPA: 3.95

CHALLENGE ME

I love challenges and the pursuit of knowledge.

Technical and engineering expertise
I can create a high quality custom replication of the Hugging Face Transformers library in a single dev day using just PyTorch.

Research and experimentation expertise
I thrive on difficult projects with high pressure and tight deadlines. Under time and hardware constraints, I will produce the best possible model.