Vaibhav Kumar

web.cs.ucla.edu/~vaibhavk vaibhavk@cs.ucla.edu github.com/TimeTraveller-San +1 (424) 402-7223

Education

- **University of California, Los Angeles** — Sep 2021 – Mar 2023
  Master of Science in Computer Science; GPA 4.0/4.0 — *Los Angeles, California*
- **Delhi Technological University** — 2016 – 2020
  Bachelor of Technology in Computer Science; GPA 8.25/10. — *Delhi, India*

Skills Summary

- **Programming Languages**: Python, C++, JavaScript, Java, Bash and Solidity
- **Frameworks**: PyTorch, HuggingFace, Langchain, Numpy, Pandas, OpenAI Gym, Scikit-Learn, Jax, LightGBM.
- **Development**: Docker, Spark, Hadoop, PostgreSQL, Flask, SQLite, jQuery, Node.js, HTML, CSS.

Experience

- **Amazon** — Sunnyvale, California
  Applied Scientist Intern — *July 2022 - September 2022 (Full-time), December 2022 (Part-time)*
  - **Natural Language Processing**: Created a lightening-fast production-ready controlled text generation framework and performed robust large scale human annotation studies on Amazon Mechanical Turk for its evaluation. Currently **submitted at the ACL 2023** conference and under internal review for application to Alexa's Web Based Question Answering (WBQA) production modules.
  - **Impact**: Achieved state of the art latency (less than +0.01s to base model) for controlled text generation with minimal loss in linguistic quality.
  - **Development**: Performed experiments for LLM training and inference on two 8 V100 GPU EC2 Instance.
- **University of California, Los Angeles** — Los Angeles, California
  Graduate Teaching Assistant — *January 2022 - Present*
  - **Teaching**: Conducted discussion lectures and held office hours for "Python with Applications I" and "Principles of Java Language with Applications" courses. Evaluated class material and solved numerous bugs.
- **UnitedHealth Group (Optum)** — Bangalore, India (Remote)
  Associate Data Scientist — *Aug 2020 – Aug 2021*
  - **Machine Learning**: Designed unsupervised models for automated intent classification and integrated them with the UHG customer care chat support. Achieved a weighted F1 score of 0.88 – evaluated over an internally human-annotated dataset carefully labelled by a diverse group of customer representatives.
  - **Impact**: Reduced the average case-assignment time from over 10 seconds (manual) to less than 1 second.
  - **Development**: Wrote cron job schedules using Bash, Apache Spark and pyspark-ml to fetch, aggregate and train machine learning models on the customer chat archive database with interactive visualizations.

Projects

- **Bayes Optimal Neural Network Pruning using Relational Graph Properties**: Using Bayes Optimization over relational graph properties of neural network connections to learn optimal neural network pruning masks.
- **Fair Embedding Engine (FEE)**: FEE provides a framework for quantifying, mitigating and visualizing gender-bias in word embeddings using a repertoire of state-of-the-art algorithms and benchmarks.
- **yTermPlayer**: A minimal, terminal based YouTube playlist streaming program written in python. Published on PyPI with more than 20,000 total downloads (*pip install ytermplayer*).
- **Epidemic Spread Mapper**: A real-time epidemic mapping system build to use crowd-sourced data for both web and android. Utilized the past temporal and spatial data to predict future epidemic spreads using LightGBM.

Publications

- **Nurse is Closer to Woman than Surgeon? Mitigating Gender-Biased Proximities in Word Embeddings**: **Vaibhav Kumar**, Tenzin Singhay Bhotia, Vaibhav Kumar, Tanmoy Chakraborty - Published in the Transactions of the Association for Computational Linguistics (**TACL**), Volume 8, 2020 p.486-503.
- **Fair Embedding Engine: A Library for Analyzing and Mitigating Gender Bias in Word Embeddings**: **Vaibhav Kumar**, Tenzin Singhay Bhotia, Vaibhav Kumar - Accepted at the 2^{nd} Workshop for NLP Open Source Software at Empirical Methods in Natural Language Processing (**NLP-OSS, EMNLP**), 2020.

Achievements

- **Generative Dog Images:** Won my second Kaggle silver medal and levelled up to a Kaggle competitions expert.
- **Santander Customer Transaction Prediction**: Bagged a silver Kaggle medal with an AUC-ROC score of 0.91