

Ethan Kim

(617) 758-9553 | ethan_kim@college.harvard.edu | 308 Amory St, Boston, MA 02130

Github: <https://github.com/ethankim00/> | Website: <https://ethankim00.github.io/>

EXPERIENCE

CYNDX

Data Scientist

Nov '21 –

- NLP Algorithms

ORACLE

Data Science Intern

May '21 – Aug '21

- Design and Implement architecture to Incorporate SHAP value-based model interpretability into Oracle's Marketing ML platform
- Deploy explanation capability for 6+ internal models including XGBoost, Random Forest, and NLP models
- Research and Implement methods to explain Recommender Systems
- Achieve 83% explainability precision for Latent Factor Models

HARVARD AI4LIFE LAB

Aug '21-Present

Research Assistant

- ML interpretability research with professor Hima Lakkaraju

THINKCERCA

Data Science Consultant

May 21 - Present

- Fine-tune BERT models for Automatic Essay Scoring and Feedback

BGI LLC

Data Science Intern

Sep '20 – Jan '21

- Develop transfer learning and data augmentation techniques for Speech to Text Transcription using Mozilla DeepSpeech to achieve over 90% Character Accuracy Rate

HARVARD COLLEGE DATA ANALYTICS GROUP

Director of Consulting

Aug '20 - Present

- Manage 24+ Data Science consulting teams working with nonprofits, startups, and Fortune 500 companies
- Sourced clients to help student club grow 5x to 100+ members with \$130,000 in revenue
- Provide technical guidance to case teams on projects ranging from marketing analytics to using NLP to scrape financial databases to billing anomaly detection

Case Team Leader

Jan '20 – May '20

- Lead team of six on Data Science research project for a sports analytics company
- Develop win probability prediction models for soccer matches that achieved high Ranked Probability Scores with faster prediction time than existing models

Coronavirus Response Team

May '20 – June '20

- Collaborate with Brigham and Women's Hospital to create data visualizations for a high impact COVID paper (JAMA IM)

BRAINSPEC

Data Science Intern

May '20 – August '20

- Create Python data processing pipeline to analyze changes in brain spectra due to PTSD and concussive injury

Interests: Harvard Club Soccer, Harvard Square Homeless Shelter

EDUCATION

HARVARD UNIVERSITY

Bachelor's in Statistics, Secondary in Computer Science

Graduation December 2021

GPA: 3.73/4.0

SAT: 2370/2400 (V800, M800, W770)

COURSES

Deep Learning, Data Structures and Algorithms, Data Science, Time Series, Probability, Inference, Linear Models, Generalized Linear Models, Computational Biology, Parallel Computing, Natural Language Processing

SKILLS

Python, Scikitlearn, TensorFlow, Pytorch, Dask, XGBoost, SQL, R, git, PySpark, Docker, AWS

PUBLICATIONS

[Google Big Bench](#)

- Contribute to suite of NLP Benchmarks

PROJECTS

Current Projects:

- Solo paper applying wav2vec to novel audio classification problem
- Research parameter efficient techniques for few shot learning in NLP

[Involution](#)

- Pytorch implementation of Convolutional Operator alternative

[Parallelizing OpenCV for Real Time Object Tracking](#)

- Use OpenMP to Parallelize multiobject tracking across multiple cores to run in real time

[Impact of Weather on Social Distancing](#)

- Scrape county level weather data to evaluate relationship with social distancing indices
- Use LSTM models to predict rates of coronavirus spread

Detection of Cell Cycle Genes and Regulators in *S. Cerevisiae*

- Time series analysis on mRNA expression data in yeast to identify key cell cycle proteins

[Training an AI to play Incan Gold](#)

- Implement playable version of Incan Gold

Ethan Kim

(617) 758-9553 | ethan_kim@college.harvard.edu | 308 Amory St, Boston, MA 02130

Github: <https://github.com/ethankim00/> | Website: <https://ethankim00.github.io/>

- AI trained on winning moves from thousands of games achieved
150% increase in win percentage