ETHAN JOSEPH

San Francisco, CA | 650-703-5345 | eaj73@cornell.edu
https://ethanhasa.website/ | https://www.linkedin.com/in/theethanjoseph/

EDUCATION

Cornell Tech, New York, NY — Aug 2022 – May 2023
Master of Engineering in Computer Science | Merit Scholar
Relevant Coursework: Applied Data Science, Machine Learning Engineering, Applied Machine Learning, Advanced Game AI, Cryptography

Rensselaer Polytechnic Institute, Troy, NY — Aug 2019 – Dec 2021
Bachelor of Science in Computer Science | GPA: 3.90, *summa cum laude*

TECHNICAL SKILLS

Coding Languages:	Python, C++, Dart, TypeScript/JavaScript, SCSS, Rust, SQL, C
Tools & OS:	Git, GCP, AWS, Docker, Firebase, Windows, Linux, Android
Libraries & Frameworks:	PyTorch, TensorFlow, CUDA, Transformers, Flutter, Numpy, Pandas, scikit-learn

EXPERIENCE

Orbitax, ML Engineer, Remote — May 2022 – Aug 2022
- Designed a recommender system for a digital marketplace used by hundreds of tax professionals.
- Built prototype for a natural language generation system utilizing GPT-3 to automatically generate country-by-country summaries for ESG reporting.

SubstrateAI, Research Intern, Remote — Jan 2022 – May 2022
- Researched usage of neural attention mechanisms in reinforcement learning agents on visual environments, improved performance on certain Atari environments by over 200%.
- Wrote pipeline to improve training/evaluation efficiency by 5x, deployed parallel GCP compute instances to automatically train and evaluate different agent architecture configurations.

Rensselaer Polytechnic Institute, Undergraduate Researcher, Troy, NY — Jan 2020 – Dec 2021
- Developed NLP powered web-scraper to extract vector-borne disease data from thousands of outbreak reports in collaboration with NASA Goddard researchers, published and presented at IEEE Big Data Workshop.
- Authored research paper on improving tabular data-to-text generation using transformer neural networks, outperformed previous SOTA by over 10% on certain metrics.
- Curated tasks and trained baseline models for novel dataset on common-sense inference.

Rensselaer Center for Open Source, Project Lead, Troy, NY — Jan 2020 – Dec 2021
- Led a team of over 15 developers from zero full-stack mobile or web dev experience to closed beta for ***smartrider***, an open-source mobile/web app that consolidates all of RPI's transportation services into a single location.
- Overcame a myriad of technical problems created by covid, including the complete restructuring of school transportation systems.

PROJECTS

BlockDreams, (Python, Rust) — Spring 2022
Attempt at creating a deep learning smart contract: GAN generated NFT with all computation occuring purely on-chain.
- Converted a GAN architecture's codebase from PyTorch to TensorFlow and trained it 12x faster using Google Cloud TPUs.
- Used quantization techniques to reduce memory footprint of the model from 0.5GB to 1MB with similar performance.
- Wrote a library to host the model on the Solana blockchain, but gas prices make deployment unfeasible.

fakeBlock, (Python, Typescript) — Spring 2021
Developed an "adblocker" style browser extension for fake news using Tensorflow.js and React.
- Modified a transformer neural network architecture to be more "explainable" and classify millions of real and fake news articles.
- Created React frontend so users can see an explainability and statistics page when the network blocks fake news.
- Deployed as a browser extension on the chrome web store, computation occurs locally to ensure user privacy.

PUBLICATIONS

- Yao, B., Joseph, E., Lioanag, J., & Si, M. A Corpus for Commonsense Inference in the Story Cloze Test. In Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC 2022) (pp. 3500–3508).
- Joseph, E., Munasinghe, T., Tubbs, H., Bishnoi, B., & Anyamba, A. (2021, December). Scraping Unstructured Data to Explore the Relationship between Rainfall Anomalies and Vector-Borne Disease Outbreaks. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 4156-4164). IEEE.
- Joseph, E., Lioanag, J., & Si, M. (2021). Improving Data-to-Text Generation via Preserving High-Frequency Phrases and Fact-Checking. IJCoL. Italian Journal of Computational Linguistics, 7(7-1, 2), 223-244.