Aneek Barua

Machine Learning Engineer - New York City Metropolitan Area

📞 646-407-8553 ✉ anbarua@cs.stonybrook.edu 🔗 linkedin.com/in/baruaaneek ⬡ github.com/AbsolutUnit

Education

Stony Brook University   **Aug. 2019 – May 2023**
Bachelor of Science in Computer Science (AI/ML Specialization) & Proof-based Mathematics   *Stony Brook, NY*

- **Relevant Courses:** Machine Learning, Artificial Intelligence, Computer Vision, Natural Language Processing, Cloud Computing, Software Engineering, Advanced Linear Algebra, Advanced C++, Analysis of Algorithms

Full Time Work Experience

Twitter, Inc.   **May 2022 – August 2022**
Machine Learning Engineer Intern   *New York, NY*

- Identified and resolved bottlenecks in our Ads Clickthrough Rate Prediction model, improving per-request time by 76% by utilizing batching and concurrency alongside Java and Rust concurrency in our request receiver.
- Discovered and addressed the true bottleneck in the system, which was not previously identified by the team, leading to a 20% increase in the throughput of the pipeline.
- Led experimental discussions on request system optimization, pushing for a Multi-Producer Single-Consumer implementation with concurrent virtual threads, resulting in a 20% increase in the overall throughput of the system.
- Developed and implemented a statistics reporter using Java and Rust that broke down 5 key components of each request feature's lifecycle into a visible graph, identifying choke points and optimizing error detection by 50%.
- **Skills Used:** Java, Rust, Machine Learning and Deep Learning, Cloud Computing, Data Analysis and Experimentation, Team Collaboration

Work Experience

Stony Brook LUNR Lab   **August 2022 – Current**
Undergraduate AI Natural Language Research Assistant   *Stony Brook, New York*

- Developed a UI in Javascript using React in order to collect event-state and cause-effect pairs in narratives and form Event State Graphs, resulting in a 30% increase in efficiency of data collection and processing.
- Assisted with implementing few-shot learning using Python to downscale various Large Language Models such as GPT-3, and automate the process of generating Event State Graphs, reducing data generation time by 70%.
- Contributed to a research publication on modeling participant states in narratives, which is currently under review and expected to be submitted in June 2023.

ObjectSecurity LLC.   **September 2022 - February 2023**
NLP Engineer Intern   *Remote*

- Developed an n-gram and frequency-based autocorrect system achieving 90% accuracy for words both inside and outside of vocab, with automated vocab population for new words.
- Implemented a data validation system using Python Cerberus, improving data quality and reducing errors in data pipelines by 40%.

Rose Technology Ltd.   **August 2021 – May 2022**
Machine Learning Engineer Intern   *New York, NY*

- Successfully optimized Dockerfiles and a Python Flask backend for a large-scale data processing pipeline, reducing overall image size from 3 GB to 4 MB and increasing pipeline throughput by over 50%.
- Developed an OpenCV-based graph reader that converted graphs to tabular data, improving data importing, processing efficiency, and computational cost by 80%.

Projects

Pay Attention | *Python, PyTorch, Torchvision*   **April 2021**

- Implemented RNN and Transformer decoders for image caption generation on Flickr8k dataset.
- Found that the Transformer decoder with positional embeddings outperformed the RNN decoder by 10% on average, with more consistent results evaluated via BLEU score, reported in a faux publication.

Technical Skills

Languages: Python, Java, C, C++, Rust, Golang, JavaScript, SQL
Developer Tools: VSCode, Git, AWS, Google Cloud Platform, Microsoft Azure
Technologies/Frameworks: PyTorch, TensorFlow, OpenCV, Linux, Docker, Kubernetes, Pandas, NumPy, Scikit-learn