Jeremy Hsu

jeremyhsu.me | jeremyhsu@college.harvard.edu | (650)-218-4411

Education

Harvard University | B.A. in Computer Science · **GPA** 3.72 May '24

Selected Coursework Data Structures and Algorithms · Operating Systems · Database Systems · Computational Complexity Probability · Functional Programming · Design of Useful Interactive Systems · Linear Algebra and Differential Equations

Experience

Meta (Facebook) | Software Engineer Intern Menlo Park, CA · May '22 – Present

- Implemented the backend for 3D avatar sticker categories on platforms such as the Facebook news feed using Hack (PHP) and GraphQL, driving increased user engagement with avatars and helping them find their desired sticker over 23% faster.
- Designed and built a new internal tooling website using Hack, React, and Relay that enables engineers, cross-functional partners, and graphic artists to directly manage avatar stickers and categories.
- Refactored client GraphQL queries to use the new 3D sticker category backend architecture and wrote a migration script in Hack to create new 3D sticker categories based on existing 2D sticker categories.

Lime | Software Engineer Intern San Francisco, CA · May '21 – Aug '21

- Architected and developed the end-to-end refund flow, user lifecycle API, and an endpoint to fetch real-time vehicle data through QR code scanning or license plate entry for a server-side API build with Ruby on Rails and MySQL, enabling seamless MaaS integrations between Lime vehicles and public transit apps.
- Worked closely with engineers and cross-functional partners at Lime and external partners to ship the API to production and complete a successful inaugural launch in Berlin in August 2021.
- Extended the functionality of a company-wide internal administrative portal by creating React UI components that support assigning permission access-control roles and viewing mobility data feed instructions for regions and external partners.

Cisco | Software Engineer Intern San Jose, CA · May '20 – May '21

- Significantly improved developer productivity by building new Python-based internal tools, deployed using Jenkins or CLI interfaces, to query, analyze, and triage client logs for a cross-platform media SDK for Webex Meetings and Teams.
- Implemented a new data aggregation layer in the SDK's metrics ETL pipeline using Python, the Elasticsearch API, and Airflow, increasing Kibana dashboard load times by over 500%.
- Created and monitored interactive Kibana dashboards visualizing critical media quality and network performance metrics, enabling Webex engineers to rapidly identify problematic trends and issues.

Activities

Harvard Data Systems Lab | Research Assistant Cambridge, MA · Feb '22 – Present

- Advised by Prof. Stratos Idreos and Utku Sirin on designing a cascading classifier composed of lightweight CNNs built in Python and Keras, which aims to reduce training time, improve model adaptability, and address subproblem heterogeneity.
- Implemented an optimal subtask partitioning algorithm based on centroid distances with k-means clustering and benchmarked a custom ResNet-based model to determine the optimal hyperparameters for different partitions of the ImageNet dataset.

Datamatch | Algorithm Lead Cambridge, MA · Sep '19 – May '22

- Led a team of 20+ students to redesign and add improvements such as user bio sentiment analysis to Datamatch's scoring and stable machine algorithms written in C++ and Python, which has scaled to match 50,000+ students across 40+ schools.
- Improved Datamatch's pairwise compatibility score function by implementing sentence embeddings with Sentence-BERT to analyze user sentiment similarities and dynamic inverse-proportional weightings to address answer distribution polarities.

Projects

CoffeeOS C++ · ASM

CoffeeOS is a multicore x86-64 operating system that supports kernel task suspension, system calls, virtual memory, and multithreading. Additional features include wait queues, caching, prefetching, synchronization primitives such as spinlocks and futexes, and an on-disk file system supporting file extensions and directory trees.

ShuttleDB C

ShuttleDB is a persistent column-store storage engine supporting fast select-project-join queries on columns with up to 10M tuples. In addition to the parser, optimizer, and executor, ShuttleDB features internal query optimizations such as B+ tree indexing, custom TCP streaming protocols, prefetching, zone maps, and multithreaded scan sharing.

ML-Enabled Spotify Curator Flask · React · Celery · Redis · Scikit-Learn · Spotify Web API

Spotify Curator is a website that predicts whether a user will like a particular song based on empirical music data. Spotify Curator uses a random forest classifier deployed on a distributed task queue to enable non-blocking client requests.

Technical Skills

Programming Languages C · C++ · Java · Python · JavaScript · PHP · Swift

Frameworks & Technologies React · Ruby on Rails · Node.js · TensorFlow · Keras · Firebase · Elasticsearch · GraphQL · SwiftUI