

Vaed Prasad

GitHub://vaedprasad • LinkedIn://vaed
201-995-3287 • vaedprasad@gmail.com

EDUCATION

Cornell University

M.ENG. COMPUTER SCIENCE
May 2021 | Ithaca, NY
Specialization in Machine Learning
GPA: 3.91/4.00

Cornell University

B.S. COMPUTER SCIENCE
Dec 2020 | Ithaca, NY | Dean's List
Minor in Networks, Crowds, & Markets
Minor in Law & Society

Horace Mann School

Jun 2017 | Bronx, NY
President, Mathematics Team
Honors in Computer Science
Honors in Mathematics

SKILLS

LANGUAGES

Python • Java • Swift • SQL
OCaml • Julia • JavaScript
HTML/CSS • BASH • L^AT_EX

TOOLS & FRAMEWORKS

Docker • Kubernetes • AWS
GCP • Git • Snowflake • DataDog
CUDA • PyTorch • OpenCV
scikit-learn • SciPy • NumPy

COURSEWORK

Deep Learning*
Computer Vision*
Machine Learning*
Learning with Big Data*
Natural Language Processing*
Programming Languages and Logics*
Artificial Intelligence*
Database Systems*
Algorithms
Data Structures
Functional Programming
Object Oriented Programming
Discrete Structures
Operating Systems
Embedded Systems
Digital Logic & Computer Org.
UNIX Tools & Scripting
Probability for Random Signals

* Denotes Graduate-Level Course

HONORS

8VC Engineering Fellow
Congressional Award Gold Medal
Eagle Scout with Gold Palm

EXPERIENCE

Snorkel AI | SOFTWARE ENGINEER, MACHINE LEARNING

Sep 2023 – Present | Foundation Models Team | Redwood City, CA

- Led the integration of **Freeform LLMs** in Snorkel Flow for enhanced question-answering, information extraction, and sequence tagging (**Python**)
- Architected an extensible **Foundation Model** (FM) configuration interface, facilitating customer integration of custom in-house for prompting (**Python**)
- Integrated **Retrieval-Augmented Generation** (RAG) using embeddings for large-document chunking, reducing inference latency and costs (**Python**)
- Awarded **1st Place** in Hackathon (voted by founders), **Engineering Excellence Award** (voted by founders), and **Best Technical Demo Award** (voted by the R&D organization)

PathAI | SOFTWARE ENGINEER, MACHINE LEARNING

Jun 2022 – Aug 2023 | MLOps Team | Boston, MA

- Served as **Technical Lead**, enabling orchestration and ensuring precision of machine learning workflows leveraging CUDA across various GPU models: RTX8000, A40, P2 (K80), and P3 (V100) (**Python, SQL, Kubernetes, Docker**)
- Served as **Machine Learning Frames Steward** (resident-expert), leading research and development initiatives for frames-based model evaluation pipelines (**Python**)
- Leveraged multiprocessing to accelerate consensus-based frames model evaluation pipelines by **>97%** and integrated these post-inference evaluation pipelines into real-time model training validation metrics (**Python**)
- Served as a **Certified Readability Reviewer** to improve the code style guide, amplify pithy documentation efforts, and enforce rigorous testing across code reviews

Bubble | SOFTWARE ENGINEER INTERN

Jan 2022 – Apr 2022 | Language Theory Team | New York, NY

- Led development to enable API workflows with multiple parameters for page and backend custom events (**JavaScript**)

PathAI | SOFTWARE ENGINEER CO-OP, MACHINE LEARNING

Jun 2021 – Dec 2021 | Clinical Pathology Research & AutoML Teams | Boston, MA

- Led research and development of frames-based sampling algorithm, leveraging hierarchical clustering to improve annotation diversity for active learning (**Python**)
- Spearheaded Bayesian consensus-based model evaluation pipeline to score cell predictions and rate pathologist annotation accuracy with the EM algorithm (**Python**)

Amazon Lab126 | SOFTWARE DEVELOPMENT ENGINEER INTERN

May 2020 – Aug 2020 | Alexa AI Conversations Team | Sunnyvale, CA

- Overhauled Alexa's Dialog Simulator to support expanding Alexa's deep conversational model for future third-party extensions (**Java, Python**)
- Implemented Golden Goal Sampler using pre-trained word embeddings to generate synthetic annotated conversations for supervised training (**Java, Python**)

Audible | SOFTWARE DEVELOPMENT ENGINEER INTERN

Jun 2019 – Aug 2019 | Core iOS Team | Newark, NJ

- Developed Wishlist module, dynamically presenting user's product wishlist on-demand with queuing functionality: **deployed in version 3.10** (**Swift**)
- Built Listening Time module, calculating user's daily, weekly, and monthly audio-book listening time: **deployed in version 3.11** (**Swift**)

Nuance Communications | SPEECH & DATA SCIENCE INTERN

May 2018 – Jun 2018; Jul 2017 – Aug 2017 | Acquired by Microsoft | Mahwah, NJ

- Performed hyperparameter optimization for transfer routing algorithms, reducing misroutes in caller-agent matching (**Python**)
- Developed dynamic grammars to enable IVR technologies for phone systems (**Python**)