PRANAV DULEPET

+1(925) 997-0461 ◇ San Ramon, CA
pdulepet@umd.edu ◇ linkedin.com/in/pranavdulepet ◇ pranavdulepet.github.io

EDUCATION

B.S./M.S. in Computer Science - ML, University of Maryland, College Park — Expected May 2026
Honors: Computer Science Honors, Dean's List, QUEST Honors Program
Courses: Deep Learning, Machine Learning, AI, Algorithms, Data Structures, Object-Oriented Programming I/II, Data Science, Computer Vision, Linear Algebra, Calculus I/II, Probability & Statistics, Discrete Structures

SKILLS

Languages/Technologies Python, Java, C, Swift, Ruby, JavaScript, Git, AWS, OpenAI Gym & GPT, LLMs
Libraries/Frameworks TensorFlow, PyTorch, Keras, FastAPI, Pandas, MongoDB, Firebase, React, MSFT Z3

EXPERIENCE

Software Developer Intern (ML), ***Amazon*** — Jun 2024 - Aug 2024
- Developed personalized routine suggestions pipeline using customer-Alexa interaction data
- Built data pre-processing framework with PySpark for over 60TB of interaction data
- Prompted and fine-tuned Claude 3 Sonnet through AWS Bedrock to generate structured and cohesive outputs from customer interaction data with a acceptance rate of 87%

Software Engineer Intern, ***Fidelity Investments*** — Jun 2023 - Aug 2023
- Built LinkedIn-like MyNetwork recommendation engine for internal Fidelity app for 80k employees
- Achieved recommendations with 98% satisfaction rate during initial user testing
- Used Python, PyTorch, DGL, Swift to build a custom Graph Neural Network to train and inference
- Identified bugs/improvements in internal app and increased code coverage by 50%

Undergraduate Researcher, ***PIRL (PI: Professor Ramani Duraiswami)*** — Jan 2023 - Present
- Developed a factorable attention mechanism reducing transformers' complexity to $O(N)$, inspired by fast multipole and Gauss transform methods (view on arxiv)
- Ensured this streamlined process still captures complete data relationships, avoiding data loss often seen with similar methods
- Previously worked with Swift, LiDAR, Autonomous Reinforcement Learning simulations

Machine Learning Intern, ***Capital One*** — Jan 2023 - May 2023
- Implemented NMSLIB similarity search frameworks on financial graph embeddings as part of the Enterprise Graph Services Team to detect transaction fraud
- Applied to samples of up to 5 million in size with high-dimensional outputting >90 recall (success rate)
- Tested framework with Merchant-Account data resulting in similar recall

Software Engineer Intern, ***Evozyne*** — Jun 2022 - Aug 2022
- Developed SMT solvers (Z3) in Python to decrease runtime of modeling the Gene Synthesis process by 5x while maintaining precision
- Visualized Gene Synthesis data to determine where the current model lacked efficiency and precision using ligation matrices, statistical fidelity, and Seaborn plots
- Explored SMT's potential use cases in Gene Regulation Networks, Reversing Genomes, Protein Folding

PROJECTS

agora. *Large Language Models, LangChain, Python, SwiftUI, Swift, AWS, MongoDB, Rest APIs*
Developed iOS app and fine-tuned LLM to provide personalized and affordable meals for students. Adapted Stable Diffusion to generate visuals. Integrated Amazon Fresh and Kroger API to buy ingredients. (website link)

College RO *Swift, SwiftUI, Python, Node.js, Rest APIs, MongoDB, AWS, Google/Firebase Analytics*
Launched CollegeRO on the App Store helping college students find research opportunities, reaching a peak of 1.5k app units. (app link)

PUBLICATIONS

- FAST: Factorizable Attention for Speeding up Transformers
- The Prompt Report: A Systematic Survey of Prompting Techniques