Vineeth Vajipey

184 Claremont Ave, #3E, New York, NY, 10027
(848) 213-7010 vv2372@columbia.edu vineethvajipey.github.io

Education

Columbia University **Sep' 2023 – May' 2025**
Master of Science in Computer Science, Thesis *New York, New York*

University of Maryland **Sep' 2019 – May' 2023**
Bachelor of Science in Computer Engineering, GPA: 3.97/4.0 *College Park, Maryland*

Relevant Coursework

- Advanced Data Structures
- Algorithms
- Linear Algebra
- Deep Learning
- Applied Machine Learning
- Natural Language Processing
- Intro to EE (TA)
- Electric Circuits I (TA)
- Reinforcement Learning (Teacher)

Experience

Speech and Dialog Lab | *HuggingFace, Pytorch, Tensorflow, Sklearn* **Aug '23 – Present**
NLP Researcher *New York, New York*

- Conducting NLP and LLM research at Columbia University advised by Dr. Zhou Yu and Qingyang Wu.
- Fine tuned LLaMA and GPT-2 with 70k user-shared conversations (GPT-4) to build "mini-Vicuña". Cleaned and formatted Share-GPT dataset. Utilized HuggingFace and FastChat for training, serving, and evaluation.
- Built full stack system to collect high quality task oriented data. Data used to benchmark language model API calling capabilities. Training position based, interface driven language models to replace unreliable API calls.

RAAS Lab | *C++, NVIDIA, OpenCV, PyTorch, ROS* **Dec '21 – May '23**
Robotics Researcher *College Park, Maryland*

- Researched and conducted extensive physical experiments for automatic bridge inspection with DJI Matrice 600 drone, Velodyne LiDAR, and onboard PC. Advised by Dr. Pratap Tokekar (RAAS Lab @ UMD).
- Implemented novel LiDAR based online path-planning algorithm, performed LiDAR-Camera calibration and sensor fusion, and developed defect detection computer vision model.
- Published research paper to ICUAS 2023, titled *'GATSBI: An Online GTSP-Based Algorithm for Targeted Surface Bridge Inspection*, and competed in the Unmanned Aerial Vehicle (UAV) Competition.

Capital One | *Python, Tensorflow, OpenCV, Microsoft Azure* **May '20 – Dec '20**
Machine Learning Researcher *College Park, Maryland*

- Researched 3D object detection and tracking from a birds-eye-view and presented findings at annual University of Maryland symposium as part of FIRE - Capital One: Machine Learning program.
- Developed ML and IoT based tools to digitized laboratory experiments impacted by Covid-19. Utilized by 500 students over 3 semesters. Advised by Dr. Romel Gomez and Dr. Raymond Tu at UMD.
- Presented products to the Philippine-American Academy of Science and Engineering, International Digital Twin Consortium, and at the Microsoft Azure-IoT Conference.

Projects

CharactAR: An Augmented Reality Customizable Real-time Chatbot | *Python, React, Swift* **Sep '23 – Sep '23**

- Designed and developed AR personal dialogue based conversational avatar using OpenAI GPT+Shap-e API, 8th Wall, 11 Labs, and Google Cloud Platform within 36 hrs. Details: https://devpost.com/software/charactar
- Won grand prize (1st place) at the Fall 2023 PennApps Hackathon with over 400 participants.

Automatic Lane Following and Obstacle Avoidance w/ Imitation Learning | *Python, C++* **May '23 – Aug '23**

- Engineered full-stack custom racecar with Traxxas chasis, mounted NVIDIA MiniPC, VESC, and ZED stereo camera.
- Implemented fully autonomous driving on dynamic environments and obstacle configurations, and tested impact of speed, data-size, sensor suite (**rgb**, *depth*, **rgb** + *depth*) on performance.
- Developed web-server UI to perform DAgger and Dataset Augmentation for improved data collection and training.

Lyric Generation enhanced by Symbolic Music using Transformers | *Python, PyTorch, AWS* **Jan '23 – Mar '23**

- Designed, implemented, and trained a model to generate song lyrics based on lyric prompts and corresponding MIDI files by combining and adapting GPT-2 and MusicBERT architectures.
- Constructed training dataset by combining music files from Lakh MIDI dataset with song lyrics from Genius API.
- Evaluated model output against GPT-2 and GPT-2 finetuned on the same lyrics dataset.

Technical Skills

Languages: Python (PyTorch, TensorFlow, Pandas, NumPy, Sklearn), C, C++, R, MATLAB, JavaScript, SQL, Swift
Developer Tools: Git, AWS, Azure, PostgreSQL
Awards: PennApps: *1st Place* (Fall 2023), *4th Place* (Fall 2021), *Best Hack for Resilience* (Fall 2019)