

# Rajan Vivek

+1 602 463 6531 ♦ rvivek@alumni.stanford.edu ♦ rvivek3.github.io ♦ /in/rajanvivek717

## Education

---

<b>Stanford University</b>	2022 – 2024
MS Computer Science: Artificial Intelligence Specialization	GPA: 4.06
<b>Georgia Institute of Technology</b>	2018 – 2022
BS Electrical Engineering: Signal Processing and Data Science Coursework Focus	GPA: 3.96

## Technical Skills

---

<b>Languages</b>	Python, C++, Java, MATLAB, SQL, C, HTML
<b>Machine Learning</b>	PyTorch, Tensorflow, scikit-learn (sklearn), numpy, pandas, Deep Graph Library, ROS
<b>Cloud Computing</b>	AWS Lambda, S3, EC2, API Gateway, DynamoDB, Domino, SageMaker, MapReduce
<b>Relevant Courses</b>	Machine Learning, NLP, Foundation Models, Speech Processing, Statistics, Decision Making

## Experience

---

<b>Contextual AI</b>	Fall 2023
<i>Research Intern</i>	
Ran 200+ experiments for LM (draft) model distillation for speculative decoding, achieving 2-2.6X inference speed up. Implemented semantic entropy uncertainty estimation for LLMs and productionized with LLM serving platform.	
<b>Stanford NLP Group</b>	Fall 2022 – Winter 2023
<i>NLP Researcher</i>	
Designed representative data selection technique for efficient language model evaluation, leading to EACL publication. Studied benchmark hill-climbing during LM pretraining, data-relatedness metrics, zero-shot vs. fine-tuned performance.	
<b>Scale AI</b>	Summer 2023
<i>Machine Learning Research Engineering Intern</i>	
Ran 100+ experiments to design novel spacio-temporal Q-former for video foundation model, surpassing InstructBLIP, VideoLLaMA, & MPlugOwl at causal and temporal video question answering (on NextQA). Used AWS SageMaker.	
<b>JPMorgan Chase Asset Management</b>	Summer 2022
<i>Deep Learning Intern</i>	
Developed and productionized a transformer-based entity recognition model for invoice processing. Designed custom data augmentation techniques, improving model performance by 11% (F1-score). Used AWS Lambda, S3, DynamoDB.	
<b>Lockheed Martin Missiles and Fire Control: Applied Research</b>	Summer 2021 & 2020
<i>Deep Learning Intern</i>	
2020: Wrote image processing scripts for end-to-end classification pipeline and experimented with CNN architectures. 2021: Performed 60+ experiments with embedding algorithms for large graph data and segmentation for 3D point clouds.	

## Publications

---

<i>Anchor Points: Benchmarking Models with Much Fewer Examples</i>	EACL 2024 Main (Long Paper)
<b>Rajan Vivek</b> , Kawin Ethayarajh, Diyi Yang, Douwe Kiela	
<i>Explainable Activity Recognition for Smart Home Systems</i>	ACM Transactions on Interactive Intelligence 2023
Devleena Das, Yasutaka Nishimura, <b>Rajan Vivek</b> , Naoto Takeda, Sean T. Fish, Thomas Ploetz, Sonia Chernova	

## Projects and Awards

---

**CS330 Outstanding Project Award (2023):** Top of ~100 teams for “Synthetic Data Generation for Few-Shot Learning”

**NSF Graduate Research Fellowship Honorable Mention (2022):** National recognition for achievements in undergraduate research in explainable AI, as well as K-12 STEM outreach and robotics curriculum development.

**HackGT 2019:** Received two 1<sup>st</sup> place awards and 2<sup>nd</sup> overall (of 250 teams) for *Smooth.io*: A voice assistant-connected food scale and iOS app that calculates and visualizes nutritional content of ingredients in real-time. Used REST API, Arduino, C.

**Opportunity Research Scholars 2019:** 2<sup>nd</sup> overall (of 25 teams) for “Robust Deep Learning-Based Motion Planner”