Vinay **Samuel**

Portfolio | Github | vinaysamuel2003 | vsamuel@andrew.cmu.edu | 713-705-6174

Education

Aug 2021 – Jun 2025 | **Carnegie Mellon University** | **Pittsburgh, USA**
Bachelor of Science in Statistics and Machine Learning

Relevant Coursework: (PhD) Large Language Models [**Python**, **PyTorch**], (PhD) Introduction to Deep Learning [**Python**, **PyTorch**], (PhD) Convex Optimization, (PhD) Advanced Natural Language Processing, Introduction to Machine Learning [**Python**, **PyTorch**], AI Problem Solving and Representation

Research Experience

Jun 2024 – Present | **Stanford NLP** | **Palo Alto, CA**
Researcher | Primary Advisor: Prof. Diyi Yang

- Collaborated on the development of Co-Gym, a novel interactive platform facilitating **human-AI collaboration** across multiple tasks, such as travel planning, related-works generation, and tabular data analysis.
- Architected and implemented a specialized environment for collaborative Related-Works generation, enabling dynamic interaction between human researchers and **AI agents** to synthesize literature reviews based on research queries
- Developed and executed a comprehensive evaluation framework to assess human-AI collaborative effectiveness in literature review tasks, encompassing both simulated and real-user experimental conditions

May 2024 – Present | **Carnegie Mellon University LTI** | **Pittsburgh, PA**
Lead Researcher | Primary Advisor: Prof. Daphne Ippolito

- Designed a novel framework for precise length control of total words in LM generation through **supervised fine-tuning** and **model editing**. [First Author]
- Curated a length constraint dataset with 50k instruction tuning samples each containing a precise word count to facilitate future research and benchmarking of length control techniques.
- Currently expanding framework application to multiple attributes such as average word length and sentiment analysis.

Mar 2024 – Aug 2024 | **Princeton NLP Group** | **Princeton, NJ**
Lead Researcher | Primary Advisor: Prof. Karthik Narasimhan

- Established the pioneering dynamic **evaluation framework**, PersonaGym, for **persona agents** in **Large Language Models (LLMs)**, and developed a novel metric, PersonaScore, to quantify LLMs' role-playing capabilities, demonstrating exceptional innovation and research acumen. [First Author]
- Spearheaded the comprehensive implementation of PersonaGym, including critical design decisions and advanced **prompt engineering** techniques to optimize performance across various components, exhibiting strong leadership and project management skills.
- Evaluated 6 open-source and proprietary **Large Language Models (LLMs)** across 200 diverse personas, encompassing 10,000 evaluation questions, and benchmarked their average PersonaScore using the PersonaGym framework, showcasing rigorous analytical capabilities.

Apr 2024 – Aug 2024 | **University of Illinois Chicago NLP Group** | **Chicago, IL**
Lead Researcher

- Led the comprehensive evaluation of 5 state-of-the-art **data contamination detection** methods across 4 large language models (LLMs) on 8 diverse and challenging datasets. [First Author]
- Developed a novel pilot contamination detection technique that leveraged dataset sequence order, highlighting contamination detection limitations during instruction fine-tuning with answer augmentation.
- Identified critical inconsistencies and limitations across multiple SOTA contamination detection methods, providing a unified analysis framework for future contamination research.
- Spearheaded the development of a contamination oracle to simulate and analyze contamination scenarios, revealing the urgent need for robust detection methods in fine-tuned LLMs.

Dec 2023 – Feb 2024 | **University of Illinois Chicago NLP Group** | **Chicago, IL**
Researcher | Primary Advisor: Prof. Cornelia Caragea

- Constructed first ever publicly available dataset for multimodal Implicit Attribute Value Extraction, termed ImplicitAVE, with 68K training data and 1.6K testing data across 5 different domains.
- Spearheaded the benchmarking of 5 state of the art **multimodal large language models (MLLMs)** on ImplicitAVE. Prompt engineered and conducted hyperparameter search for optimal performance of multimodal large language models on ImplicitAVE, showcasing leadership in experimental design and execution. [ACL Findings Paper]

Apr 2023 – Aug 2023 | **Independent Research Group** | **Remote**
Lead Researcher | Primary Advisor: Aman Chadha

- Leveraged **GPT-4** to generate 200k+ QA pairs, boosting diversity of training data and for smaller Bert-like LMs [**Hugging Face**] to generalize to out-of-domain questions competing with gold standard human-annotated data sets such as SQuAD, demonstrating initiative and creative problem-solving. [First Author]
- Augmented 3 low resource datasets using GPT4 to increase exact match by 5% - 27% and f1 score by 2% - 15% on test datasets of 3 well-known low resource datasets. [ACL SRW Paper]

Oct 2021 – Jun 2022 | **Xu Lab at Carnegie Mellon University** | **Pittsburgh, PA**
Computer Vision Intern | Primary Advisor: Prof. Min Xu

- Fine-tuned the state-of-the-art imaging models using **Tensorflow** to make tile-level mutational signature predictions and the **LSTM** to make patient-level predictions from the intermediate layer embeddings of the **ResNet CNN model**.
- Leveraged statistical modeling and cross validation to determine optimal hyperparameters for model.
- Achieved 1.2% increase in previous accuracy which would help clinicians to assess and detect a wide variety of conditions such as smoking and UV radiation exposure.

Select Research Publications

S=In Submission, C=Conference, W=Workshop | Complete List at Google Scholar

[S] **PersonaGym: Evaluating Persona Agents and LLMs** [Paper] [arXiv] [website]
Vinay Samuel, Henry Peng Zou, Yue Zhou, ... , Ameet Deshpande, Karthik Narasimhan, Vishvak Murahari
Currently Under Review at AAAI 2025 **[Under Review]**

[C] **Towards Data Contamination Detection for Modern Large Language Models: Limitations, Inconsistencies, and Oracle Challenges** [Paper] [arXiv]
Vinay Samuel, Yue Zhou, Henry Peng Zou
Coling 2025 **[Coling 2025]**

[C] **ImplicitAVE: An Open-Source Dataset and Multimodal LLMs Benchmark for Implicit Attribute Value Extraction** [Paper] [arXiv] [Code]
Henry Peng Zou, **Vinay Samuel**, Yue Zhou, Weizhi Zhang, Liancheng Fang, Zihe Song, Philip S. Yu, Cornelia Caragea
ACL Findings 2024 (Meta Score: 4, Soundness: 4/4/4, Overall: 4/3.5/3.5) **[ACL 2024]**

[W] **Can LLMs Augment Low-Resource Reading Comprehension Datasets? Opportunities and Challenges** [Paper]
Vinay Samuel, Houda Aynaou, Arijit Ghosh Chowdhury, Karthik Venkat Ramanan, Aman Chadha
ACL Student Research Workshop 2024 **[ACL SRW 2024]**

Awards and Honors

Carnegie Mellon University Deans List Spring 2024

Carnegie Mellon University Deans List High Honors Fall 2023

Carnegie Mellon University Deans List Spring 2022

Carnegie Mellon University Deans List Fall 2021

Teaching

Convex Optimization (PhD), Carnegie Mellon University *Teaching Assistant* Aug'24 - Present

- Design weekly problems and solutions for homework and quizzes along with exam questions for 40 graduate students.
- Conduct weekly office hours to instruct and guide students on assignments and course material.

Skills

Languages and Libraries: Python, Pytorch, Tensorflow, Hugging Face Transformers, Numpy, Pandas, OpenCV, Open-LLM, Keras, C++, Java, R

Research: Personalization in LLMs, Question Answering (QA), Multimodal Attribute Value Extraction, LM Controlability, Data Contamination Detection, LM Agents, Model Evaluation, NLP, CUDA, LLM Fine-tuning, Prompt Engineering, Deep Learning