

Vashisth Tiwari

✉ vashistt@cs.cmu.edu | 🌐 vashistht.github.io/ | **in** [vashistht](https://www.linkedin.com/in/vashistht)

EDUCATION

Carnegie Mellon University (Pittsburgh, PA) Aug 2023 – Dec 2024
Master of Science in Artificial Intelligence Engineering GPA: 4.0/4.0

- **Relevant Courses:** Deep Learning Systems, Advanced NLP, Deep Learning, Machine Learning, AI Systems, Stochastic Processes

University of Rochester (Rochester, NY) Aug 2019 – May 2023
Bachelor of Science in Physics, Bachelor of Arts in Mathematics GPA: 3.97/4.0

- **Relevant Courses:** Modern Statistics & Exploration (Grad), Data Structures & Algorithms, Probability, Honors Real Analysis, Honors Linear Algebra, Honors Multi-Dimensional Calculus, Differential Equations

RESEARCH EXPERIENCE

Active Learning For Synthetic Data Generation Pittsburgh, PA
Research Assistant (w/ Prof. Emma Strubell, Carnegie Mellon University) Aug 2024 – Present

- Designing a responsive-feedback driven framework where teacher models iteratively refine synthetic data generation based on student model's performance and learning outcomes.

Efficient Machine Learning Pittsburgh, PA
Research Assistant (w/ Prof. Beidi Chen, Carnegie Mellon University) Feb 2024 – Aug 2024

- Showcased how speculative decoding can mitigate the tradeoff between throughput and latency in LM inference.
- Implemented weight pruning, attention sparsity, and activation sparsity techniques for drafting in self-speculation
- Evaluated compressed models' effectiveness through acceptance rate and speedup metrics in speculative decoding
- Achieved up to a 2x speedup over autoregressive baseline for LLaMA-3-8B inference at high batch sizes (≥ 128) through self-speculation and sparse key-value optimizations.

ML for Dark Energy Spectroscopic Instrument Rochester, NY
Research Assistant (w/ Prof. Segev Benzvi, University of Rochester) Jan 2020 – May 2021

- Designed multi-class CNNs for spectral data with TensorFlow, scikit-learn to find galaxies with supernovae
- Enhanced network performance by applying noise-removal techniques to preprocess spectral data
- Achieved 95%+ accuracy and high precision for supernovae classification tasks in the DESI data pipeline

Blok Lab (Quantum Computing with Superconducting Qubits) Rochester, NY
Research Assistant (w/ Prof. Machiel Blok, University of Rochester) Sep 2021 – May 2023

- Characterized the noise in a quantum computer by comparing the probability distributions of the lab data with the expected distributions from quantum and statistical physics
- Utilized neural networks to classify qubit readouts of different energy levels to reduce readout errors in the setup

Polymath Research Experience for Undergraduates (REU) Online
Research Assistant (w/ Prof. Steven Miller, Williams College) Jul 2020 – Aug 2020

- Contributed two proofs related to the bounds on the length of the Zeckendorf Game, a number theory project
- Verified these conjectures for large numbers using Mathematica and Python scripts

WORK EXPERIENCE

Mana Finance Corporation Hillsborough, CA
Quantitative Research Intern (Mentor: Max Novendstern, David Kaufman) May 2022 – Aug 2022

- Utilized statistical techniques to analyze stock price distributions and quantify investment risk
- Developed ML models using Facebook Prophet for assessing expected yields on potential investments
- Prototyped a tool demonstrating direct tracking of Ethereum blockchain data on UniSwap

Los Alamos National Laboratory Los Alamos, NM
Research Intern (Mentor: Dr. Malcolm Boshier) Jun 2021 – Aug 2021

- Modeled complex quantum system using Python and utilized Mathematica numerical differential equation solvers
- Discovered optimal laser pulse parameters through high-dimensional optimization
- Improved the system performance by 5% beyond the current state-of-the-art pulse parameters through optimized pulses

PROJECTS

- Deep Learning Systems** | CMU Aug 2024—Dec 2024
- Implementing a PyTorch-like Deep Learning library with auto differentiation, optimizer, and GPU support module
 - Adding auto-diff support for Fourier analysis and linear algebra operators like inverses, eigenvalue decompositions
- YapperJay: Aligning ASR with Human Preferences** | CMU Oct 2024 – Present
- Applying RLHF techniques to enhance ASR transcription quality, aligning with natural human preference using Direct Preference Optimization (DPO).
 - Utilizing Google FLEURS dataset with paired examples and fine-tuned a pre-trained ASR model from SpeechLM-Toolkit to improve case sensitivity, punctuation, and discourse marker handling.
- Consumer Hardware Pruning with Preserved Reasoning** | CMU Mar 2024 – May 2024
- Improved upon Bonsai (forward pass only structured LLM pruning, compatible with consumer hardware) in math-reasoning tasks
 - Demonstrated the scope for using novel task-aware pruning metric to better retain the reasoning abilities than the standard perplexity baseline
- End-to-End NLP System Building** | CMU Feb 2024 – Mar 2024
- Engineered a Retrieval Augmented Generation (RAG) based chatbot on CMU utilizing webpages and semantic scholar data
 - Implemented core RAG components: LangChain embedder, Faiss+ColBERT retriever, and reader using open-source LLMs

TEACHING EXPERIENCE

- Carnegie Mellon University** | *Language Technology Institute* Fall 2023
- **Courses Taught:** Advanced Natural Language Processing (11-711)
 - **Responsibilities:**
 - * Mentored students on term research projects; held weekly office hours to support assignments and address conceptual questions.
 - * Designed and graded assignments, including developing a baseline Retrieval-Augmented Generation (RAG) system for students to benchmark their RAG performance against.
- University of Rochester** | *Dept. of Physics & Astronomy, Dept. of Computer Science* 2020 – 2023
- **Courses Taught:**
 - * Quantum Theory (Spring 2023)
 - * Advanced Electromagnetism (Fall 2022)
 - * Honors Waves and Modern Physics (Spring 2021)
 - * Introduction to Programming in Python (Fall 2020)
 - **Responsibilities:**
 - * Led weekly workshops with practice problems and concept reviews
 - * Mentored 30+ students through office hours and tutoring sessions; designed and graded homework and exams

AWARDS & HONORS

- Undergraduate Teaching Award**, Dept. of Physics & Astronomy 2023
- Harry W. Fulbright Prize**, University of Rochester (*2 students in the graduating class for excellence in experimental physics*) 2023
- Phi Beta Kappa**, National Honor Society 2023
- Semi-Finalist, Rhodes Scholarship**, Indian Consulate 2022
- Sigma Pi Sigma**, National Physics Honor Society 2022
- Physics Honors Prize**, University of Rochester (*Highest grade in first two years of honors physics*) 2021
- Summer Research Grant**, University of Rochester (*\$1500 grant, 66 recipients*) 2020
- Next Genius Scholar**, University of Rochester (*Full scholarship for undergrad*) 2019

PUBLICATIONS & PRESENTATIONS

MACHINE LEARNING

MagicDec: Breaking the Latency-Throughput Tradeoff for Long Context Generation 2024
Chen J*, **Tiwari V***, Sadhukhan R* et al. (*Equal contribution). *Accepted to ECCV Efficient Foundation Model Workshop*

PHYSICS

Improved Bragg splitting of Bose-Einstein condensates into high-order momenta wave-packets 2023
Uzun C, Pandey S, **Tiwari V**, Krzyzanowska K, Boshier M. *American Physical Society Division of Atomic, Molecular and Optical Physics (DAMOP)*

High-fidelity splitting of Bose-Einstein condensates into high-order momentum states 2022
Uzun C, Pandey S, **Tiwari V**, Krzyzanowska K, Boshier M. *American Physical Society DAMOP*

Using Machine Learning to Develop a Transient Identification Pipeline for DESI 2021
Wasserman A, **Tiwari V**, BenZvi S.
📍 *Co-Presented at the 237th Meeting of the American Astronomical Society*

MATHEMATICS (ALPHABETICAL AUTHOR LIST)

Bounds on Zeckendorf Games 2022
Cusenza A., Dunkelberg A., Huffman K., Ke D., McClatchey M., Miller S. J., Mizgerd C., **Tiwari V.**, Ye J., and Zheng X.. *Fibonacci Quarterly*, 60 (2022), no. 1, 57–71

📍 *Co-presented the work at the Young Mathematicians Conference with Carl Ye and Kevin Ke*

Winning Strategy for Multiplayer and Multialliance Zeckendorf Games 2021
Cusenza A., Dunkelberg A., Huffman K., Ke D., Kleber D., Miller S. J., Mizgerd C., **Tiwari V.**, Ye J., and Zheng X.. *Fibonacci Quarterly*, 59 (2021), no. 4, 308–318

📍 *Co-presented the work at the Young Mathematicians Conference and UConn Mathematics Conference*

Extending Zeckendorf's Theorem to a Non-constant Recurrence 2020
Boldyriev E., Cusenza A., Dai L., Ding P., Dunkelberg A., Haviland J., Huffman K., Ke D., Kleber D., Kuretski J., Lentfer J., Luo T., Miller S. J., Mizgerd C., **Tiwari V.**, Ye J., Zhang Y., Zheng X., and Zhu W. *Fibonacci Quarterly*, 58 (2020), no. 5
📍 *Presented at International Conference on Fibonacci Numbers*

SERVICE & LEADERSHIP

ECE Department, Carnegie Mellon University Pittsburgh, PA
Peer Mentor Aug 2023 – Present

- Mentor new international students in navigating academic and personal life at CMU
- Connect mentees with relevant resources and support systems within the university

Society of Physics Students (SPS) Rochester, NY
President (Previously Secretary) Aug 2021 – May 2023

- Initiated STEM education program through interactive DIY activities for Rochester City School District students, focusing on low-income, Black, and Hispanic students
- Led chapter to win **outstanding chapter award** (top 15%) for community building and physics outreach
- Organized tutoring program serving 300+ students in introductory physics classes

Mentorship Rochester, NY

- “How to Intern at National Labs?”: Summer Opportunities Talk, Society of Asian Students and Engineers (2021)
- \LaTeX Workshop, “ \LaTeX : A Brief Introduction and the Essentials”, Society of Physics Students (2020, 2021)

SKILLS & INTERESTS

Languages: Python (Expert), C++ (Intermediate), Java (Intermediate), Bash (Proficient), Mathematica

ML/AI: PyTorch, HuggingFace, TensorFlow, LangChain, Faiss, CUDA

Data/Systems: NumPy, Pandas, Scikit-Learn, Spark, Kafka, AWS (EC2, S3, SageMaker), Docker

Tools: Git, Linux, Jupyter, LaTeX

Interests: Competitive Badminton (University Team), Photography, Hiking, Frisbee