Siddharth Parekh

Pittsburgh, PA | P: +1 332-248-8513 | sid@parekh.co.in

EDUCATION

CARNEGIE MELLON UNIVERSITY Pittsburgh, PA
Master of Science in Computer Science (Expected Aug 2026)
Upcoming Coursework: Advanced Natural Language Processing, Computational Game Solving, Deep Reinforcement Learning

Bachelor of Science in Computer Science, Minor in Machine Learning (May 2025)
Graduated with University Honors and SCS College Honors
Cumulative GPA: 3.8/4.0; Dean's List: Fall 2021, Fall 2022, Spring 2023, Fall 2024, Spring 2025
Selected Coursework: Learning and Game Theory, Natural Language Processing, Deep Learning, Convex Optimization, Probabilistic Graphical Models, Fixed Income Markets, Database Systems, Data Structures and Algorithms

EXPERIENCE

Student Researcher Jan 2024 – Present

- SCS Undergraduate Honors Thesis: Investigated efficient post-training methods to enhance correctness in Large Language Models (LLMs) for Question Answering through combinations of methods like:
  - Equilibrium Ranking: A decoding algorithm used to reconcile generative and discriminative querying
  - Expert Persona Implementation: Leveraging specialized knowledge domains by having the model adopt various expert personas and comparing them to existing methods like Few-Shot prompting and Chain-of-Thought reasoning.
- Advanced visually rich document understanding with Prof. Carolyn Rose and Armineh Nourbaksh:
  - Developed robust evaluation metrics for document visual question answering datasets like DocVQA, DUDE, InfographicVQA, MultiPage-DocVQA that focus on incorporating grounding with correctness. (**NAACL Findings 2025**).
  - Contributed to a Graph Neural Network-based model for Form Processing, achieving performance comparable to state-of the-art models in Key-Information and Relation Extraction (LayoutLMV3, GeoLayoutLM) with 30% fewer parameters (**EMNLP Findings 2024**).

Language Technologies Institute, CMU Pittsburgh, PA
Research Intern Jun 2023 – Aug 2023

- Led research on numeracy in Large Language Models (LLMs) under Professor Carolyn Rose, focusing on financial document parsing and question answering tasks using datasets like FinQA, TAT-QA, and PACIFIC.
- Improved baseline models' accuracy by 3-5% through integration of code-generation LLMs (T5, CodeT5), enhancing their ability to process numerical data in complex financial contexts.

Miko AI Mumbai, India
AI Intern - NLP Jun 2022 – Aug 2022
Streamlined the multilingual personality module of the Miko robot, enhancing its language processing capabilities:

- Benchmarked Helsinki-NLP machine translation models and Google Cloud Translate API using cosine similarity of SIEBERT-generated sentence embeddings, optimizing cost-efficiency without compromising performance.
- Developed a classification system for multilingual question answering using feed-forward neural networks, achieving linear speedup over traditional vector search methods.

ACTIVITIES

Tartan Student Fund
Associate Portfolio Manager Aug 2022 – Dec 2023
Developed an automated system for providing analysts with updated news and statistics on equities to assist TSF, CMU's student managed long-equity investment fund with approximately $100k in management, with weekly coverage of our holdings.

Teaching Assistant Jan 2023 – May 2023

- Served as the sole undergraduate teaching assistant for CMU's premier NLP course, 11-411/611 Natural Language Processing, led by Prof. David Mortensen and Prof. Lori Levin.
- Managed a diverse cohort of undergraduate and graduate students, providing instruction, resolving queries, and mentoring teams throughout a semester-long project focused on question answering and question generation.

ADDITIONAL

Programming: C/C++, Python, Java, Javascript, SQL
ML Frameworks: NumPy, Scikit-Learn, PyTorch, TensorFlow, Transformers, HuggingFace
Miscellaneous: Docker, AWS, Google Cloud Platform, React, MS Office, Adobe CC, LaTeX
Languages: Fluent in English, Gujarati; Conversational Proficiency in Hindi; Elementary Proficiency in Spanish