Tanmay Parekh

tparekh97@gmail.com | tparekh@cs.ucla.edu | tanmayparekh.github.io
tanmayparekh | TanmayParekh | tparekh97
Los Angeles, California

Education

University of California Los Angeles (UCLA), Los Angeles, USA — *Sep '21 - Ongoing*
Doctor of Philosophy (Ph.D.) in Computer Science — GPA: 4.0/4.0
Advisors: Prof. Nanyun Peng and Prof. Kai-Wei Chang

Carnegie Mellon University (CMU), Pittsburgh, USA — *Aug '19 - Jul '21*
Masters of Science (MS) in Language Technologies — GPA: 4.0/4.0
Advisors: Prof. Alan Black, Prof. Yulia Tsvetkov, and Prof. Graham Neubig

Indian Institute of Technology Bombay (IITB), Mumbai, India — *Jul '14 - May '18*
Bachelor of Technology (B.Tech.) with Honors in Computer Science and Engineering — GPA: 9.37/10.0
Advisors: Prof. Preethi Jyothi and Prof. Shivaram Kalyanakrishnan

Research Interests

LLM Agents: Efficient Planning and Search, Memory Management for long-horizon tasks
Post-training LLMs: Efficient Reasoning, Synthetic Data Generation
AI for Applications: Code Generation, Question Answering, Multilinguality, Long-context Understanding and Generation

Industry Experience

Bloomberg AI — *Jun '25 - Oct '25*
Data Science Intern, Managers: Dr. Yunmo Chen and Dr. Srivas Prasad — New York, USA
- Proposed a novel **software-engineering inspired agentic framework** for complex code generation
- Introduced parallel planning and **single-step multi-path exploration** in the agentic workflow to optimize the Pareto-optimality for performance-latency
- Achieved the **state-of-the-art** (SOTA) on text-to-SQL benchmark SPIDER 2.0. Paper under submission

Meta GenAI — *Jun '24 - Oct '24*
Research Scientist Intern, Managers: Pradyot Prakash and Dr. Denis Savenkov — Bellevue, USA
- Developed DyPlan, a **dynamic strategy planning and self-correcting framework** that routes queries to the optimal reasoning path (e.g., CoT, advanced planning, or RAG) based on input difficulty
- **Fine-tuned and aligned LLMs** on synthetically generated data to calibrate their difficulty assessment
- Demonstrated improvements in performance and efficiency on QA. Published at **NAACL 2025**

Amazon Alexa — *Jun '22 - Sep '22*
Applied Scientist Intern, Managers: Dr. Christopher Hench and Dr. Jing Huang — Sunnyvale, USA
- Designed an alignment algorithm to **align semantic concepts (e.g. negation) in multilingual LMs**

Amazon Machine Learning — *Jul '18 - Jun '19*
Applied Scientist, Managers: Sachin Farfade and Dr. Subhajit Sanyal — Bengaluru, India
- Developed an attribute extraction model using distant supervision and semi-supervised self-training

Selected Academic Research

Divergent-Convergent LLM Reasoning for Zero-shot Classification *[EMNLP '25]*
- Proposed a two-stage LLM reasoning comprising: (1) **open-ended unconstrained (divergent) reasoning** to boost the recall and (2) **task-specific grounded (convergent) reasoning** to improve precision
- Achieved SOTA zero-shot results on six event extraction datasets across 5 diverse technical domains

Event-centric Evaluation of Personalized News Narratives *[Ongoing]*
- Studying LLMs' capability to write personalized news narratives based on the selection and organization of atomic events from a large pool of real-world facts

LLM-based Synthetic Data Generation for Low-resource domains *[EMNLP '25]*

- Designed a **multi-agent synthetic data generation framework** to extract domain-specific signals from unsupervised text, inversely generate supervised data ($Y \rightarrow X$), and verify the data quality.
- **Fine-tuned LLMs and SLMs** on our synthetic data and evaluated on downstream event extraction task
- Demonstrated the efficacy in news, epidemiology, and biomedicine domains across three languages

Label Projection for Multilingual Data Generation [Best Paper Nomination] *[NAACL '24]*

- Utilized instruction-tuned LLMs to perform **contextual machine translation of labels** for generating zero-shot cross-lingual data for structured prediction tasks
- Demonstrated improved cross-lingual results on Information Extraction tasks across 50+ languages

PUBLICATIONS

[18] **Tanmay Parekh**, Ella Hofmann-Coyle, Shuyi Wang, Sachith Sri Ram Kothur, Srivas Prasad, Yunmo Chen. **PExA: Parallel Exploration Agent for Complex Text-to-SQL**. Under review at *ACL Rolling Review (ARR)*, January 2026.

[17] **Tanmay Parekh**, Kartik Mehta, Ninareh Mehrabi, Kai-Wei Chang, and Nanyun Peng. **DiCoRe: Enhancing Zero-shot Event Detection via Divergent-Convergent LLM Reasoning**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

[16] **Tanmay Parekh**, Yuxuan Dong, Lucas Bandarkar, Artin Kim, I-Hung Hsu, Kai-Wei Chang, and Nanyun Peng. **SNaRe: Domain-aware Data Generation for Low-Resource Event Detection**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

[15] **Tanmay Parekh**, Pradyot Prakash, Alexander Radovic, Akshay Shekher, and Denis Savenkov. **Dynamic strategy planning for efficient question answering with large language models**. In *Findings of the Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2025.

[14] **Tanmay Parekh**, Jeffrey Kwan, Jiarui Yu, Sparsh Johri, Hyosang Ahn, Sreya Muppalla, Kai-Wei Chang, Wei Wang, and Nanyun Peng. **SPEED++: A multilingual event extraction framework for epidemic prediction and preparedness**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[13] Ashima Suvarna, Xiao Liu, **Tanmay Parekh**, Kai-Wei Chang, and Nanyun Peng. **QUDSELECT: Selective decoding for questions under discussion parsing**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[12] Kuan-Hao Huang, I-Hung Hsu, **Tanmay Parekh**, Zhiyu Xie, Zixuan Zhang, Premkumar Natarajan, Kai-Wei Chang, Nanyun Peng, and Heng Ji. **TextEE: Benchmark, Reevaluation, Reflections, and Future Challenges in Event Extraction**. In *Findings of the Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

[11] **Tanmay Parekh**, I-Hung Hsu, Kuan-Hao Huang, Kai-Wei Chang, and Nanyun Peng. **Contextual label projection for cross-lingual structured prediction**. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2024. [Best Paper Nomination]

[10] **Tanmay Parekh**, Anh Mac, Jiarui Yu, Yuxuan Dong, Syed Shahriar, Bonnie Liu, Eric Yang, Kuan-Hao Huang, Wei Wang, Nanyun Peng, and Kai-Wei Chang. **Event detection from social media for epidemic prediction**. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2024.

[9] **Tanmay Parekh**, I-Hung Hsu, Kuan-Hao Huang, Kai-Wei Chang, and Nanyun Peng. **GENEVA: Benchmarking generalizability for event argument extraction with hundreds of event types and argument roles**. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[8] Amrith Setlur*, Aman Madaan*, **Tanmay Parekh***, Yiming Yang, and Alan W. Black. **Towards using heterogeneous relation graphs for end-to-end tts**. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021.

[7] **Tanmay Parekh**, Emily Ahn, Yulia Tsvetkov, and Alan W. Black. **Understanding linguistic accommodation in code-switched human-machine dialogues**. In *Proceedings of the 24th Conference on Computational Natural Language Learning (CoNLL)*, 2020.

[6] Fanglin Chen, Ta-Chung Chi, Shiyang Lyu, Jianchen Gong, **Tanmay Parekh**, Rishabh Joshi, Anant Kaushik, and Alexander Rudnicky. **Tartan: A two-tiered dialog framework for multi-domain social chitchat**. In *Proceedings of the Alexa prize proceedings*, 2020.

[5] Aman Madaan*, Amrith Setlur*, **Tanmay Parekh***, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W. Black, and Shrimai Prabhumoye. **Politeness transfer: A tag and generate approach**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[4] **Tanmay Parekh**, Sachin Farfade and Nikhil Rasiwasia. **Automatic and Accurate Attribute Extraction for E-Commerce**. In *Proceedings of the Amazon Machine Learning Conference (AMLC)*, 2019.

[3] Saurabh Garg*, **Tanmay Parekh***, and Preethi Jyothi. **Code-switched language models using dual rnns and same-source pretraining**. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2018.

[2] **Tanmay Parekh**, and Preethi Jyothi. **Language Modelling for Code-Switched Text**. Undergraduate Thesis at *Indian Institute of Technology Bombay (IITB)*, 2018.

[1] Saurabh Garg, **Tanmay Parekh**, and Preethi Jyothi. **Dual language models for code switched speech recognition**. In *Proceedings of Interspeech (19th Annual Conference of ISCA)*, 2018.

*Indicates Equal Contribution

Scholastic Achievements and Awards

- Awarded three highly competitive PhD fellowships: **Bloomberg Ph.D. Data Science Fellowship** (2025-26), **Amazon Science Ph.D. Fellowship** (2024-25), and **UCLA Computer Science Fellowship** (2021-22)
- Represented CMU at Alexa Socialbot Challenge 2020 and reached the Semifinals
- Received **ISCA Student Grant** at Interspeech '18

Professional Services

- Awarded as an **Outstanding Reviewer** at EMNLP 2025
- Served as a **Reviewer** for ACL (2021-2025), NAACL (2022-2025), EMNLP (2022-2025), EACL (2023-2025), SRW (2023-2025).
- Served as the **Program Chair** for the Socal NLP Symposium 2023
- Founding member of the UCLA NLP Seminar Series

Teaching Services

Teaching Assistant, University of California Los Angeles (UCLA) *2022-2023*
- Introduction to Machine Learning (CS 146)
- Introduction to Natural Language Processing (CS 162)
- Natural Language Processing - Graduate Division (CS 263)

Teaching Assistant, Carnegie Mellon University (CMU) *2020-2021*
- Multilingual Natural Language Processing (11-737)
- Introduction to Speech Processing (11-492)
- Advanced Speech Processing (11-692)

Teaching Assistant, Indian Institute of Technology Bombay (IITB) *2015-2018*
- Introduction to Machine Learning (CS 419)
- Computer Architecture Theory and Lab (CS 305 and 341)
- Linear Algebra (MA 106)
- Calculus (MA 105)

Skills

Programming Languages: Python, PyTorch, Tensorflow, MySQL, Git, LaTeX, C++, Matlab, R

Languages: Hindi (native), English (fluent), Marathi (intermediate), Spanish (basic)