Sanket Vaibhav Mehta

sanketvmehta@google.com | sanketvmehta.github.io

Research Interests

- Deep Learning, Machine Learning, Natural Language Processing, Optimization
- **Focus:** Mid-Training/ Continual/ Lifelong Learning and Reinforcement Learning for LLMs
- **Doctoral Thesis:** Efficient Lifelong Learning in Deep Neural Networks: Optimizing Architecture, Training, and Data (Thesis committee: Emma Strubell, William W. Cohen, Aditi Raghunathan, Dani Yogatama)

Education

Carnegie Mellon University, Pittsburgh, USA — July 2020 - November 2023
Ph.D. in Language Technologies, LTI, SCS — CGPA: **4.07/4.33**
Advisor: Emma Strubell

Carnegie Mellon University, Pittsburgh, USA — August 2019 - February 2020
Ph.D. in Language Technologies, LTI, SCS — CGPA: **4.08/4.33**
Advisor: Jaime Carbonell

Carnegie Mellon University, Pittsburgh, USA — August 2017 - August 2019
Master of Science in Language Technologies, LTI, SCS — CGPA: **4.13/4.33**
Advisors: Jaime Carbonell & Barnabás Póczos

Indian Institute of Technology Roorkee, Roorkee, India — July 2011 - June 2015
Bachelor of Technology in Computer Science and Engineering — CGPA: **9.67/10.0**
President's Gold Medalist — Institute Rank: **1** (out of 987 students)

Research Experience

Google DeepMind, Mountain View, USA — November 2023 - Present
Senior Research Scientist

- Developing advanced training paradigms for LLMs, leveraging mid-training optimization, continual/lifelong learning, and reinforcement learning for enhanced efficiency, adaptation, and scaling.

Carnegie Mellon University, Pittsburgh, USA — August 2017 - November 2023
Graduate Research Assistant

- My doctoral thesis focuses on designing efficient lifelong learning systems that alleviate catastrophic forgetting of previously learned knowledge and facilitate continuous learning of new tasks. Inspired by biological learning processes and the progress in deep learning, my work injects appropriate inductive biases into the three main components of data-driven machine learning: model (architecture and initialization), training (objective and optimization), and data (limited labeled and unlabeled).
- Earlier work focused on developing machine learning models for part demand forecasting and part price prediction (as a part of Boeing / Carnegie Mellon Aerospace Data Analytics Lab academic research initiative).

Google Research, Mountain View, USA — May 2022 - November 2022
Research Intern/ Student Researcher (Hosts: Yi Tay, Jai Gupta)

- We introduce DSI++, a continual learning challenge for DSI that requires incrementally adding documents to the model, and propose an approach that focuses on training dynamics and data-based regularization to enable it (EMNLP 2023).

Google AI, Pittsburgh, USA (Remote) — June 2021 - October 2021
Research Intern/ Student Researcher (Hosts: Yi Tay, Jinfeng Rao)

- We systematically study the problem of compositional generalization for data-to-text generation and propose a generic BLEURT-based self-training approach to improve the model's generalization capabilities (ACL 2022).
- Contributed to a suite of 107 NLP tasks, where we show that massively multi-task pre-training improves downstream performance on NLP tasks, overcoming negative transfer between tasks while fine-tuning (ICLR 2022).

Adobe Research, Bangalore, India — June 2015 - July 2017
Member of Research Staff (Team Lead: Shriram Revankar, P. Anandan)

- Worked on designing algorithms for generating data-driven geo-fences to assist Adobe's digital marketing business and prevent inadvertent information disclosures by auto-tagging security policies.
- Transferred several technologies to Adobe Analytics (US Patents 9,838,843 and 11,756,058) and Adobe Experience Manager (US Patents 10,102,191 and 10,783,262).

Technical Report and Doctoral Thesis

1. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities
   Gemini Team, Google. (2025)
2. Efficient Lifelong Learning in Deep Neural Networks: Optimizing Architecture, Training, and Data
   Sanket Vaibhav Mehta
   Carnegie Mellon University. Doctoral Thesis. (2024)

Selected Publications

1. BIG-Bench Extra Hard
   Mehran Kazemi, Bahare Fatemi, Hritik Bansal, John Palowitch, Chrysovalantis Anastasiou, Sanket Vaibhav Mehta, Lalit K. Jain, Virginia Aglietti, Disha Jindal, Peter Chen, Nishanth Dikkala, Gladys Tyen, Xin Liu, Uri Shalit, Silvia Chiappa, Kate Olszewska, Yi Tay, Vinh Q. Tran, Quoc V. Le, Orhan Firat
   Annual Conference of the Association for Computational Linguistics (ACL 2025)
2. Expert Routing with Synthetic Data for Domain Incremental Learning
   Yewon Byun, Sanket Vaibhav Mehta, Saurabh Garg, Emma Strubell, Michael Oberst, Bryan Wilder, Zachary C. Lipton
   Transactions on Machine Learning Research (TMLR 2025)
3. DSI++: Updating Transformer Memory with New Documents
   Sanket Vaibhav Mehta, Jai Gupta, Yi Tay, Mostafa Dehghani, Vinh Q. Tran, Jinfeng Rao, Marc Najork, Emma Strubell, Donald Metzler
   Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)
4. An Empirical Investigation of the Role of Pre-training in Lifelong Learning
   Sanket Vaibhav Mehta, Darshan Patil, Sarath Chandar, Emma Strubell
   Journal of Machine Learning Research (JMLR 2023)
5. Making Scalable Meta Learning Practical
   Sang Keun Choe, Sanket Vaibhav Mehta, Hwijeen Ahn, Willie Neiswanger, Pengtao Xie, Emma Strubell, Eric Xing
   Conference on Neural Information Processing Systems (NeurIPS 2023)
6. Train Flat, Then Compress: Sharpness-Aware Minimization Learns More Compressible Models
   Clara Na, Sanket Vaibhav Mehta, Emma Strubell
   Findings of the Association for Computational Linguistics (EMNLP 2022)
7. Improving Compositional Generalization with Self-Training for Data-to-Text Generation
   Sanket Vaibhav Mehta, Jinfeng Rao, Yi Tay, Mihir Kale, Ankur P. Parikh, Emma Strubell
   Annual Conference of the Association for Computational Linguistics (ACL 2022)
8. ExT5: Towards Extreme Multi-Task Scaling for Transfer Learning
   Vamsi Aribandi, Yi Tay, Tal Schuster, Jinfeng Rao, Huaixiu Steven Zheng, Sanket Vaibhav Mehta, Honglei Zhuang, Vinh Q. Tran, Dara Bahri, Jianmo Ni, Jai Gupta, Kai Hui, Sebastian Ruder, Donald Metzler
   International Conference on Learning Representations (ICLR 2022)
9. Efficient Meta Lifelong-Learning with Limited Memory
   Sanket Vaibhav Mehta*, Zirui Wang*, Barnabás Póczos, Jaime Carbonell
   Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)
10. Learning Rhyming Constraints using Structured Adversaries
    Harsh Jhamtani, Sanket Vaibhav Mehta, Jaime Carbonell, Taylor Berg-Kirkpatrick
    Conference on Empirical Methods in Natural Language Processing (EMNLP 2019)

11. Gradient-based Inference for Networks with Output Constraints
Jay-Yoon Lee, Sanket Vaibhav Mehta, Michael Wick, Jean-Baptiste Tristan, Jaime Carbonell
AAAI Conference on Artificial Intelligence (AAAI 2019)

12. Towards Semi-Supervised Learning for Deep Semantic Role Labeling
Sanket Vaibhav Mehta*, Jay-Yoon Lee*, Jaime Carbonell
Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)

Issued Patents

1. Generating data-driven geo-fences (US 9,838,843)
2. Propagation of changes in master content to variant content (US 10,102,191)
3. Digital document update (US 10,489,498)
4. Tagging documents with security policies (US 10,783,262)
5. Digital document update using static and transient tags (US 10,846,466)
6. Tenant-side detection, classification, and mitigation of noisy-neighbor-induced performance degradation (US 11,086,646)
7. Intelligent customer journey mining and mapping (US 11,756,058)

Relevant Courses (CMU)

Introduction to Machine Learning (10-701), Algorithms for NLP (11-711), Neural Networks for NLP (11-747), Structured Prediction for Language (11-763), Deep Reinforcement Learning and Control (10-703), Multimodal Machine Learning (11-777), Deep Learning (10-707), Human Language for AI (11-724), Convex Optimization (10-725)

Accolades

- Department Research Fellowship, LTI, SCS (August 2017 - November 2023).
- Highlighted Reviewer of ICLR 2022.
- Spotlight at the ICML Workshop on Theory and Foundations of Continual Learning 2021.
- Recipient of **President's Gold Medal**, **Dr. A. N. Khosla Medal** and **Smt. Shashi Krishna Medal** for the Session 2014-15 for obtaining the highest CGPA amongst the B.Tech/ B.Arch/ IDD/ Int. M.Sc./ Int. M.Tech passing out students.
- Recipient of **Kathail Family Annual Excellence Award** and **Rakesh Agrawal Annual Excellence Award** presented by IIT Roorkee Heritage Foundation in 4^{th} year and 2^{nd} year of B.Tech in Computer Science and Engineering for outstanding Academic, Co-Curricular and Extra-Curricular achievements respectively.
- Recipient of **Certificate Of Trust Prize- Mr. Rai Singh Jain & Mrs. Shakuntla Devi Jain** presented by IIT Roorkee for the year 2014 and 2012 for The Student Obtaining Highest CGPA in B.Tech 3^{rd} year and 1^{st} year.

Responsibilities

- **Reviewer:** NeurIPS (2021, 2022, 2023, 2024), ICLR (2021, 2022, 2023, 2024, 2026), ICML (2023, 2024, 2025), EMNLP (2019, 2020, 2021, 2022, 2023, 2025), ACL (2019, 2020, 2022, 2023, 2025), AAAI (2021)
- **Research Mentor:** Clara Na (Ph.D. at CMU), Jared Fernandez (Ph.D. at CMU), Emily Byun (Ph.D. at CMU)
- **Student Mentor:** Mingkai Deng (Ph.D. at CMU), Jimin Sun (MS at CMU), Hwijeen Ahn (MS at CMU)
- **Teaching Assistant:** (1) Algorithms for NLP (11-711), (2) Artificial Intelligence (15-681A/IITP-01) where responsibilities included conducting recitations, holding office hours, creating and grading assignments and exams.
- **Internship Mentor:** Mentored students at Adobe Research over the summer of 2015, 2016, and 2017. This led to a paper at IMWUT 2017 and three issued patents – US 10,489,498, US 10,846,466, and US 11,086,646.
- **Founding Chair, IIT Roorkee ACM Student Chapter:** Presided over all the meetings of the chapter and of its Executive Council while serving as chair for IIT Roorkee ACM Student Chapter during the session 2014-15.
- **Ambassador for ACM:** Served as a campus ambassador for ACM at IIT Roorkee during the session 2014-15.

* denotes equal contribution