Suraj Tripathi

LinkedIn
Google Scholar
Website

Email : surajtripathi.iitd@gmail.com
Mobile : +1-412-954-8105

Education

- **Carnegie Mellon University** — Pittsburgh, PA
  Masters in Language Technologies; Fully-funded research master's in NLP/ML, GPA: 4.14/4.33 — *Aug 2021 - Present*
  Courses: Advanced NLP, Multilingual NLP, Multimodal Machine Learning, Question Answering, Computational Ethics for NLP
- **Indian Institute of Technology** — Delhi, India
  Masters in Computer Science and Engineering; GPA: 8.39/10.0 — *Aug 2015 - Aug 2017*
  Courses: Artificial Intelligence, Machine Learning, Numerical Algorithms, Probabilistic Graphical Models, Advanced Data Structures
- **Jamia Millia Islamia** — Delhi, India
  Bachelors in Computer Engineering; GPA: 9.31/10.0 — *July 2011 - July 2015*
  Courses: Data mining, Computer Graphics, Database System, Operating System, Compiler Design, Computer Architecture, Mathematics

Skills and Interests

- **Skills**: Natural Language Processing, Machine Learning, Prompt Engineering, Deep Learning, PyTorch, Python
- **Interests**: Large Language Models (LLMs), Parameter Efficient Learning, Multilingual NLP, Question Answering, Dialogue Systems, Information Extraction, Multimodal Machine Learning

Peer-Reviewed Publications

- **PRO-CS : An Instance-Based Prompt Composition Technique for Code-Switched Tasks, EMNLP'22,** [pdf]
  ***Suraj Tripathi**, Srijan Bansal, Sumit Agarwal, Teruko Mitamura, Eric Nyberg*
- **R3 : Refined Retriever-Reader pipeline for Multidoc2dial, DialDoc@ACL'22,** [pdf]
  ***Suraj Tripathi**, Srijan Bansal, Sumit Agarwal, Sireesh Gururaja, Aditya Veerubhotla, Ritam Dutt, Teruko Mitamura, Eric Nyberg*
- **Zero-shot cross-lingual open domain question answering, MIA@NAACL'22,** [pdf]
  *Sumit Agarwal, **Suraj Tripathi**, Teruko Mitamura, Carolyn Penstein Rose*
- **Input-conditioned convolution filters for feature learning, MoMM'20,** [pdf]
  ***Suraj Tripathi**, Saurabh Tripathi, Abhay Kumar, Chirag Singh*
- **Emoception: An Inception Inspired Efficient Speech Emotion Recognition Network, ASRU'19,** [pdf]
  *Chirag Singh, **Suraj Tripathi**, Abhay Kumar, Ajay Nagar, Promod Yenigalla*
- **Deep Learning Techniques for Humor Detection in Hindi-English Code-Mixed Tweets, WASSA@ACL'19,** [pdf]
  ***Suraj Tripathi**, Sushmitha Reddy Sane, Koushik Reddy Sane, Radhika Mamidi*
- **Speech Emotion Recognition Using Spectrogram & Phoneme Embedding, INTERSPEECH'18,** [pdf]
  *Promod Yenigalla, Abhay Kumar, **Suraj Tripathi**, Chirag Singh, Sibsambhu Kar, Jithendra Vepa*

Research and Work Experience

- **Amazon, AWS AI Rekognition** — Seattle, WA
  *Applied Scientist Intern, Advisors: **Nikolaos Pappas and Daniele Bonadiman*** — *May 2023 - Aug 2023*
  - **Cross-lingual Retrieval Augmented Generation (RAG)**: Working on Large Language Models (LLMs) based query transformation and reranking for improving cross-lingual passage retrieval.
- **Carnegie Mellon University** — Pittsburgh, PA
  *Graduate Research Assistant, Advisor: **Prof. Teruko Mitamura*** — *Aug 2021 - May 2023*
  - **Knowledge-directed Artificial Intelligence Reasoning Over Schemas (KAIROS)**: Working on DARPA-funded project KAIROS. Investigated event grounding across a set of extracted elements from multiple documents given its temporal context and various other attributes like entities, relations, etc.
  - **Summarization**: Devised effective approaches to learn an intermediate plan to ground the generation of abstractive summaries. Specifically, we prepend (or prompt) source documents with ordered sequences of entities mentioned in the summary. Paper submission is in progress.
- **Carnegie Mellon University** — Pittsburgh, PA
  *Graduate Research Assistant, Advisor: **Prof. Graham Neubig*** — *Aug 2022 - Dec 2022*
  - **Machine Translation**: Proposed a training paradigm that makes use of a non-deterministic distribution and assigns probability masses to various candidate translations based on their quality. This paradigm incorporates a contrastive loss defined over candidate translations produced by pre-trained translation models. Also explored improving the predictive performance of translation quality metrics like COMET for low-resource languages using uncertainty-aware approaches.

- **Cross-lingual Open Domain Question Answering** Pittsburgh, PA
  Graduate Research Assistant, Advisor: ***Prof. Teruko Mitamura and Prof. Carolyn Penstein Rose*** *Jan 2022 - July 2022*
  - **QA**: The Proposed method employs a passage reranker, the Fusion-in-Decoder technique for generation, and a wiki data entity-based post-processing system to tackle the inability to generate entities across all languages. Improved F1 and EM metrics by 3 and 4.6 points respectively compared to the baseline CORA model on XOR-TyDi dataset. Ranked 1st in the constrained setup for dev and 2nd in the test setting in MIA-Shared task submission.
- **Document-grounded Dialogue and Conversational Question Answering** Pittsburgh, PA
  Graduate Research Assistant, Advisor: ***Prof. Teruko Mitamura and Prof. Eric Nyberg*** *Jan 2022 - July 2022*
  - **QA**: MultiDoc2Dial is a conversational QA task that requires generating responses to user queries. Our proposed approach employs sparse representations for passage retrieval, a passage re-ranker, the fusion-in-decoder architecture for generation, and a curriculum learning training paradigm. Our approach shows a 12-point improvement in BLEU score compared to the baseline RAG model.
- **Answering Visual Questions through Representation Learning** Pittsburgh, PA
  Graduate Research Assistant, Course Project, Instructor: ***Prof. Louis-Philippe Morency*** *Aug 2021 - Dec 2021*
  - **QA**: We developed a model that combines coarse-level and fine-level features, enabling semantic reasoning through a module that dynamically selects the optimal features to answer questions. Additionally, we introduced two novel tasks, Object Feature Extraction (OFE) and Scenegraph Masking (SM), by leveraging new, task-specific datasets generated from scene graphs. These tasks were incorporated into a multi-task training framework, resulting in improved performance over the baseline Vilbert:12-in-1 model on the dev set. Notably, we also explored prompt-based learning in both low and high-resource settings, further enhancing the efficacy of our proposed methodology. [pdf]
- **Samsung Research Institute** Bangalore, India
  Lead Engineer, Bixby - Samsung's Voice Assistant *Oct 2017 - July 2021*
  - **Speech Emotion Recognition**: Worked on building a low latency and low memory footprint speech emotion recognition system. Our proposed framework resulted in more than a 4% increase in overall accuracy over the existing state-of-the-art methods with 62% fewer parameters compared to the benchmark emotion recognition models. This work was published at INTERSPEECH 2018. Another paper in AffComp@IJCAI 2019.
  - **NLU Tasks**: Investigated approaches for text-based intent classification and slot tagging model to map between human commands to low-level actions to be performed on the mobile device. Awarded Samsung Citizen and Excellence Award in 2018 and 2019 for outstanding research contributions and technical excellence.
- **Indian Institute of Technology** Delhi, India
  Graduate Research Assistant, Advisor: ***Prof Jayadeva*** *Aug 2015 - Aug 2017*
  - **Efficient Neural Networks**: Deep Neural Networks are both computational and memory intensive, making them difficult to deploy on mobile systems with limited hardware resources. Introduced a novel loss function to achieve sparsity by minimizing a convex upper bound on the Vapnik-Chervonenkis (VC) dimension. Also, analyzed the effectiveness of our proposed loss function in combination with techniques like quantization and pruning. [pdf]

Internship

- **Trexquant Investment LP** Remote
  Alpha Researcher *Mar 2017 - Sept 2017*
  - **Quantitative Research**: My primary role was to generate and execute profitable trading strategies by conducting thorough research and analysis of financial data and implementing market-neutral, medium-frequency quantitative trading strategies. Investigated and implemented recent academic research and applied machine learning techniques to alpha discovery and portfolio construction. Also, collaborated closely with portfolio managers, traders, and other research teams to refine investment strategies, and presented my findings and recommendations to senior management.

Teaching Assistant

- **IIT Delhi**: COL 774, Machine Learning
- **IIT Delhi**: CSL 374, Computer Networks
- **IIT Delhi**: CSL101, Introduction to Computers and Programming

Honors and Awards

- 1st on UNSEEN track for MultiDoc2Dial, DialDoc Workshop in ACL (2022), 3rd on MIA Shared Task, MIA Workshop in NAACL (2022), All India Rank 182 (99.92 percentile) in GATE - CSE 2015

Leadership and Volunteer Experience

- Graduate student mentor at LTI, CMU (2022-23). Internship mentor for undergraduate interns at Samsung Research (2019)