

SIDDHANT ARORA

Website : <https://siddhu001.github.io/>

Email : siddhantarora1806@gmail.com Phone : +1-724-680-6603

Email : siddhana@andrew.cmu.edu Github : <https://github.com/siddhu001>

RESEARCH INTERESTS

Spoken Language Understanding, Speech Processing, Interpretability, Robustness

EDUCATION

Carnegie Mellon University, Pittsburgh

Ph.D. Language Technology (Funded with Fellowship)
School of Computer Science.

Aug. 2022 - Aug. 2025 (Expected)

Adviser: Dr. Shinji Watanabe

Carnegie Mellon University, Pittsburgh

Master of Science, Language Technology (Funded with Fellowship)
School of Computer Science.

Aug. 2020 - Aug. 2022

CQPA : 4.21/4.33

Indian Institute of Technology, Delhi

B.Tech and M. Tech (Dual Degree)
Department of Computer Science and Engineering.

Jul. 2015 - Jul. 2020

B.Tech CGPA: 9.408/10 , M.Tech CGPA: 9.484/10

Department Rank: 1

PUBLICATIONS

- Token-level Sequence Labeling for Spoken Language Understanding using Compositional End-to-End Models:** [Siddhant Arora*](#), Siddharth Dalmia*, Brian Yan, Florian Metze, Alan W Black and Shinji Watanabe. EMNLP 2022 Conference [URL]
- A Study on the Integration of Pre-trained SSL, ASR, LM AND SLU models for Spoken Language Understanding:** Yifan Peng*, [Siddhant Arora*](#), Yosuke Higuchi, Yushi Ueda, Sujay Kumar, Karthik Ganesan, Siddharth Dalmia, Xuankai Chang, Shinji Watanabe. SLT 2022 Conference
- Two-Pass Low Latency End-to-End Spoken Language Understanding:** [Siddhant Arora](#), Siddharth Dalmia, Xuankai Chang, Brian Yan, Alan Black, Shinji Watanabe. Interspeech 2022 Conference [URL]
- ESPnet-SLU: Advancing Spoken Language Understanding through ESPnet:** [Siddhant Arora](#), Siddharth Dalmia, Pavel Denisov, Xuankai Chang, Yushi Ueda, Yifan Peng, Yuekai Zhang, Sujay Kumar, Karthik Ganesan, Brian Yan, Ngoc Thang Vu, Alan W Black, Shinji Watanabe. ICASSP 2022 Conference [URL]
- Explain, Edit, and Understand: Rethinking User Study Design for Evaluating Model Explanations:** [Siddhant Arora*](#), Danish Pruthi*, Norman Sadeh, William W. Cohen, Zachary C. Lipton, Graham Neubig. AAAI 2022 Conference [URL]
- Rethinking End-to-End Evaluation of Decomposable Tasks: A Case Study on Spoken Language Understanding:** [Siddhant Arora*](#), Alissa Ostapenko*, Vijay Viswanathan*, Siddharth Dalmia*, Florian Metze, Shinji Watanabe, Alan W Black. Interspeech 2021 Conference [URL]
- Blockwise Streaming Transformer for Spoken Language Understanding and Simultaneous Speech Translation:** Keqi Deng, Shinji Watanabe, Jiatong Shi, [Siddhant Arora](#). Interspeech 2022 Conference [URL]

8. **BERT Meets CTC: New Formulation of End-to-End Speech Recognition with Pre-trained Masked Language Model:** Yosuke Higuchi, Brian Yan, [Siddhant Arora](#), Tetsuji Ogawa, Tetsunori Kobayashi and Shinji Watanabe. EMNLP 2022 Conference [URL]
9. **A Tale of Two Regulatory Regimes: Automated Analysis of Semi-Parallel Privacy Policies in English and German:** [Siddhant Arora](#), Henry Hosseini, Christine Utz, Vinayshekhar Banihatti Kumar, Abhilasha Ravichander, Peter Story, Jasmine Mangat, Rex Chen, Martin Degeling, Tom Norton, Thomas Hupperich, Shomir Wilson, Norman Sadeh. LREC 2022 Conference
10. **IterefinE: Iterative KG Refinement Embeddings using Symbolic Knowledge:** [Siddhant Arora](#), Srikanta Bedathur, Maya Ramnath, Deepak Sharma. AKBC 2020 Conference [URL]
11. **Capreolus: A Toolkit for End-to-End Neural Ad Hoc Retrieval:** Andrew Yates, [Siddhant Arora](#), Xinyu Zhang, Wei Yang, Kevin Martin Jose, Jimmy Lin. WSDM 2020 Demo Track [URL]
12. **A survey on Graph Neural Network for Knowledge Graph Completion:** [Siddhant Arora](#). Arxiv 2020 [URL]
13. **Investigating Ad-Hoc Retrieval Method Selection with Features Inspired by IR Axioms :** [Siddhant Arora](#), Andrew Yates. AMIR Workshop, ECIR 2019 [URL]
14. **On Embeddings in Relational Database:** [Siddhant Arora](#) and Srikanta Bedathur. KR2ML Workshop, NeurIPS 2019 [URL]
15. **Understanding Community Rivalry on Social Media: A Case Study of Two Footballing Giants :** Sopan Khosla, [Siddhant Arora](#), Abhilash Nandy, Ankita Saxena, Anandhavelu N. HUMANIZE Workshop, IUI 2019 [URL]
16. **GATCheck: A Detailed Analysis of Graph Attention Networks:** Lovish Madaan*, [Siddhant Arora](#)*. Mining and Learning with Graph Workshop, KDD 2020 [URL]

ACADEMIC ACHIEVEMENTS

- **Top 10% of MLT class** in CMU in first 2 semesters
- **CMU Graduate Research Fellowship**
- **Institute Silver Medal** for securing Department Rank 1 in Dual Degree Program of Computer Science and Engineering at IIT Delhi
- **Merit certificate** at IIT Delhi in 7 out of 10 semesters given to only top 7% students
- All India Rank **11 (amongst 1,200,000 candidates)** in **JEE Mains 2015** and **289 (amongst 140,000 candidates)** in **IIT-JEE Advanced 2015**
- **KVPY Scholar fellowship** (300 selected amongst 600,000 candidates) in January 2014

ACADEMIC SERVICE

- Conference Reviewer in **EMNLP 2022, AAI 2022, SLT 2022**

RESEARCH EXPERIENCE

Graduate Research Assistant CMU Pittsburgh
Prof. Shinji Watanabe and Prof. Alan Black *Dec 2020 - Current*

Robust Testing and Open Source toolkit for Spoken Language Understanding

Graduate Research Assistant CMU Pittsburgh
Prof. Graham Neubig, Prof. Zachary Lipton and Prof. William Cohen *Feb. 2021 - Aug. 2021*

User Study Design for Evaluating Explanations

Graduate Research Assistant
Prof. Norman Sadeh
Multilingual Annotation for Privacy Documents

CMU Pittsburgh
Aug. 2020 - Aug. 2022

Undergraduate Research Assistant
Prof. Srikanta Bedathur

IIT Delhi
Jul 2019-Jun 2020

Learning Embeddings on Relation Databases and Knowledge Graph Refinement

Undergraduate Research Intern
Prof. Andrew Yates

MPII Saarbrucken
Jan 2019 - May 2019

Stress Testing IR Models and Open source Toolkit

Undergraduate Research Intern
Prof. Tom Gedeon

ANU Canberra
Nov 2017 - Dec 2017

Emotion Recognition using user gaze data

WORK EXPERIENCE

Research Intern
Adobe Systems
Understanding Toxic Content Online

Bangalore, KA, India
May 2018 - Jul 2018

Data Science Summer Intern
Bidgely Inc.
Energy Disaggregation for Household

Mountain View, CA, US
May 2019 - Jul 2019

TEACHING EXPERIENCE

Teaching Assistant
Prof. Norman Sadeh
Mobile and IoT Computing Services

CMU
Jan 2022- May 2022

Teaching Assistant
Prof. Shravan Kumar
MTL100 course (Calculus)

IIT Delhi
Jul 2018-Dec 2018

Teaching Assistant
Prof. Subodh Kumar
COL106 (Data Structures)

IIT Delhi
Jul 2019-Dec 2019

Teaching Assistant
Prof. Rijurekha Sen
COL380 (Parallel and Distributed Programming)

IIT Delhi
Jan 2020-May 2020

INVITED TALKS & PANELS

- “Two-Pass Low Latency End-to-End Spoken Language Understanding”. Oral Presentation, **Interspeech 2022**. September 2022.
- “Explain, Edit, and Understand: Rethinking User Study Design for Evaluating Model Explanations”. Oral Presentation, **AAAI 2022**. February 2022.
- “Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text”. **FTC Privacy Con.** June 2021.
- “Using NLP to Understand Differences in Privacy Disclosures and Protections under Different Jurisdictions”. **Cylabs Lab Tour**. October 2021.