Atharva Kulkarni

Ph.D. in Computer Science | University of Southern California

atharvajk98.github.io @ atharva.kulkarni@usc.edu github.com/atharvajk98 Google Scholar
linkedin.com/in/atharvajk x.com/atharvajk25

Research Interests: Robust, Generalizable, & Trustworthy NLP

Education

-Aug 2024 | **University of Southern California | Thomas Lord Dept. of Computer Science** — **Los Angeles, USA**
Doctor of Philosophy (Ph.D.) in Computer Science
Advisor: Prof. Swabha Swayamdipta

May 2024 Aug 2022 | **Carnegie Mellon University | Language Technologies Institute** — **Pittsburgh, PA**
Master of Science (M.S.) in Language Technologies, QPA: 4.04/4.00
Advisors: Prof. Graham Neubig & Prof. Barnabás Póczos

May 2020 Aug 2016 | **Savtribai Phule Pune University | PVG's College of Engineering and Technology** — **Pune, India**
Bachelor of Engineering (B.E.) in Computer Science, CPGA: 9.1/10

Research Experience

Aug 2024 May 2024 | **Apple Inc. | Siri & Information Intelligence (SII)** — **Seattle, WA**
Machine Learning Research Intern | Advisors: Dr. Yuan Zhang & Dr. Hong Yu
› Evaluating evaluation metrics for hallucination identification in LLMs.

May 2023 | *Machine Learning Research Intern | Advisors: Shruti Bhargava & Dr. Hong Yu*
› Synthetic conversational data generation using Large Lanugage Models. (*EACL 2024*)

May 2024 Aug 2022 | **Carnegie Mellon University | Language Technologies Institute** — **Pittsburgh, PA**
Research Assistant | Advisors: Prof. Graham Neubig & Prof. Barnabás Póczos
› Information-theoretic analysis of uncertainty quantification metrics in large language models.
› Constrained multi-task learning for improving worst-group generalization. (*TMLR 2024*)
› Factual error correction using modular LLMs and constrained decoding. (*EMNLP 2023 Findings*)

July 2022 March 2021 | **IIIT-Delhi | Laboratory for Computational Social Systems (LCS2)** [🌐] — **Delhi, India**
Predoctoral Researcher | Advisor: Prof. Tanmoy Chakraborty
› Hate speech detection under subpopulation and covariate shifts. (*SIGKDD 2023*)
› Semantic role labeling of multimodal memes using optimal transport. (*EACL 2023*)
› Claim span identification using structural constraints and domain-specific rationales. (*EMNLP 2022*)
› Sarcasm explanation generation in mult-imodal multi-party dialogues. (*ACL 2022*)

Selected Publications (∗ denotes equal contribution)

[10] **Still Not Quite There! Assessing Large Language Models for Comorbid Mental Health Diagnosis**
Amey Hengle*, Atharva Kulkarni*, Shantanu Patankar, Rashmi Gupta
[Under Review EMNLP 2024]

[9] **Multitask Learning Can Improve Worst-Group Outcomes** [PDF | code]
Atharva Kulkarni*, Lucio Dery*, Amrith Setlur, Aditi Raghunathan, Ameet Talwalkar, Graham Neubig
Transactions of Machine Learning Research **[TMLR'24]**

[8] **SynthDST: Synthetic Data is All You Need for Few-Shot Dialog State Tracking** [PDF | code]
Atharva Kulkarni, Bo-Hsiang Tseng, Joel Moniz, Dhivya Piraviperumal, Hong Yu, Shruti Bhargava.
Conference of the European Chapter of the Association for Computational Linguistics **[EACL'24]**

[7] **Counting the Bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model** [PDF | code]
Leonie Weissweiler, Valentin Hofmann, Anjali Kantharuban, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Atharva Kulkarni, Abhishek Vijayakumar, Haofei Yu, Hinrich Schuetze, Kemal Oflazer, David R Mortensen.
Conference on Empirical Methods in Natural Language Processing **[EMNLP'23]**

[6] **The Student Becomes the Master: Matching GPT3 on Scientific Factual Error Correction** [PDF | code]
Dhananjay Ashok, Atharva Kulkarni, Hai Pham, Barnabás Póczos.
Findings of the Association for Computational Linguisticss **[EMNLP'23 Findings]**

[5] **Revisiting Hate Speech Benchmarks: From Data Curation to System Deployment** [PDF | code]
Atharva Kulkarni*, Sarah Masud*, Vikram Goyal, Tanmoy Chakraborty
ACM SIGKDD Conference on Knowledge Discovery and Data Mining **[KDD'23]**

[4] **[AI for Social Good Award]** **Learning & Reasoning Multifaceted Longitudinal Data for Poverty Estimates and Livelihood Capabilities of Lagged Regions in Rural India** [PDF]
Atharva Kulkarni, Raya Das, Ravi Srivastava, Tanmoy Chakraborty
International Joint Conference on Artificial Intelligence **[IJCAI'23]**

[3] **Characterizing Entities in Harmful Memes: Who is Hero, Villain, Victim?** [PDF | code]
Shivam Sharma*, Atharva Kulkarni*, Tharun Suresh, Himanshi Mathur, Preslav Nakov, Md. Shad Akhtar, Tanmoy Chakraborty
Conference of the European Chapter of the Association for Computational Linguistics **[EACL'23]**

[2] **Empowering the Fact-checkers! Automatic Identification of Claim Spans on Twitter** [PDF | code]
Megha Sundriyal*, Atharva Kulkarni*, Vaibhav Pulastya, Md. Shad Akhtar, Tanmoy Chakraborty
Conference on Empirical Methods in Natural Language Processing **[EMNLP'22]**

[1] **When did you become so smart, oh wise one?! Sarcasm Explanation in Multi-modal Multi-party Dialogues** [PDF | code]
Shivani Kumar*, Atharva Kulkarni*, Md. Shad Akhtar, Tanmoy Chakraborty
Annual Meeting of the Association for Computational Linguistics **[ACL'22]**

Academic Service

Conference Reviewer	TMLR'24, ARR'24, EMNLP'23, ACL'23, EMNLP'22, ICON'22
Workshop Reviewer	WiNLP'24, NeurIPS'23 SSLTheoryPractice, NeurIPS'23 R0-FoMo
Wokrshop Organizer	ACL'22 CONSTRAINT Workshop
Conference Volunteer	EACL'24, EACL'23, ACL'22, ACL'21, EACL'21

Other Research Projects

Revisiting Information Theory for Uncertainty Quantification in Text Generation Jan'24 - May'24
Course: 11-910 Directed Research | Advisor: Prof. Graham Neubig

› Proposed a novel metric for uncertainty quantification in text generation – the Jensen-Shannon divergence with respect to the Zipfian distribution. The metric accounts for both the concentration of probability mass and the diversity of tokens in the generated text, offering a balanced measure of uncertainty.

On the Effectiveness of LoRA: Advantages, Pitfalls, and Misconceptions Aug'23 - Dec'23
Course: 11-910 Directed Research | Advisor: Prof. Emma Strubell

› Conducted a large scale empirical analysis of LoRA on – 1) Quantifying its training instability 2) Generalizability of its representations 3) Flatness of the loss landscape 4) Sensitivity to Hyperparameters.

$\mathcal{V}$-Usable Information based LLM Pruning Jan'23 - May'2023
Course: 11-830 Computational Ethics in NLP | Instructors: Prof. Emma Strubell & Prof. Maarten Sap

› Designed a $\mathcal{V}$-usable information based Iterative Magnitude Pruning (ITM) technique that takes into account example difficulty for LLM compression. The pruning method effectively mitigates bias, enhances worst-group robustness, and preserves top-line performance compared to naive ITM. [PDF]

Continual Learning via Winning Sub-Network Selection Jan'23 - May'23
Course: 11-707 Advanced Deep Learning | Instructor: Prof. Ruslan Salakhutdinov

› Proposed a Winning SubNetworks (WSN) based technique for worst-group generalization that identifies the most effective subnetworks for each sub-group while preserving previously chosen weights. Achieved competitive performance with group-DRO on CIFAR100. [PDF]

Training Dyamics and Entropy-Aware Mixup for Robust Text Classification Aug'22 - Dec'22
Course: 11-711 Advanced NLP | Instructor: Prof. Graham Neubig

› Developed a robust mixup technique that generates synthetic samples by interpolating pairs of high and low entropy datapoints of easy-to-learn and ambiguous examples in a label aware fashion. Reported an improvement of about 2% in f1-score on the 2 benchmarks datasets. [PDF]

Honours and Awards

1. **USC Viterbi Fellowship (2024-25)** – Awarded to selected incoming PhD student at USC, covering full tuition & stipend.
2. **CMU LTI Research Fellowship (2023-24)** – Awarded to selected MLT students at CMU, covering full tuition & stipend.
3. **Best AI for Social Good Project | IJCAI'23** [🔗] Poverty estimation using deep learning for Indian setting.
4. **Shared Task Winner | WASSA @ EACL'21** [🔗] Best system for empathy & distress detection in reaction to news stories.
5. **Shared Task Winner | TechDoFication @ ICON'20** [🔗] Best system for coarse-grained technical domain identification for documents in indigenous Indian languages.
6. **Bachelor's Thesis Award | ZS ASPIRE 2020** [🔗] 2^{nd} place amongst 33,000 teams and cash prize of 2,00,000 INR for project SmartCap – A deep learning based device for visually impaired people.

Talks

"SynthDST: Synthetic Data is All You Need for Few-Shot Dialog State Tracking"

- EACL'24 Presentation — March 2024 – St. Julian's, Malta
- Siri & Information Intelligence, Apple. Inc — Aug 2024 – Seattle, WA

"When did you become so smart, oh wise one?! Sarcasm Explanation in Multi-modal Multi-party Dialogues"

- Student Research Symposium, CMU LTI — Aug 2022 – Pittsburgh, PA

Teaching and Leadership Roles

Advanced NLP, CMU *Teaching Assistant* — Fall'23

- TA for 11711 - Advanced NLP at CMU, taught by Prof. Daniel Fried.
- Designed assignments and guided teams for semester long course projects.

Natural Language Processing, CMU *Teaching Assistant* — Spring'23

- TA for 11611 - Natural Language Processing at CMU, taught by Prof. David Mortensen.
- Designed assignments and guided teams for semester long course projects.

AI Undergrad Mentorship Program, CMU *Mentor* — Fall'22

- Mentored undergraduate students at CMU for starting out with NLP research.

Collaborative Open Research in Deep Learning (CORD.ai) *Mentor* — Fall'22

- CORD.ai is a community of research mentors from MIT, CMU, Georgia Tech, UBC, Samsung, Adobe, and Microsoft.
- Mentored students from underprivileged backgrounds on AI projects.

References

Prof. Graham Neubig — *Associate Professor, Carnegie Mellon University, USA* [🌐]

Prof. Barnabás Póczos — *Associate Professor, Carnegie Mellon University, USA* [🌐]

Prof. Tanmoy Chakraborty — *Associate Professor, IIT Delhi, India* [🌐]