

Atharva Kulkarni

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

[athrvkk.github.io](https://github.com/athrvkk) [@ atharva.kulkarni@usc.edu](mailto:atharva.kulkarni@usc.edu) github.com/athrvkk [Google Scholar](https://scholar.google.com/citations?user=athrvkk) [in linkedin.com/in/athrvkk](https://www.linkedin.com/in/athrvkk)
x.com/athrvkk [athrvkk.bsky.social](https://bsky.social/athrvkk)

Research Interests: Theoretical & Empirical Understanding of LLMs | Learning Dynamics | Trustworthy NLP

Education

May 2029 Aug 2024	University of Southern California Thomas Lord Dept. of Computer Science Doctor of Philosophy (Ph.D.) in Computer Science > Advisor: Swabha Swayamdipta > Honors: Viterbi School of Engineering Fellowship	Los Angeles, CA
May 2025 Jan 2025	University of California, Berkeley Simons Institute for the Theory of Computing Visiting Ph.D. Student Program: Special Year on LLMs and Transformers (Part 2)	Berkeley, CA
May 2024 Aug 2022	Carnegie Mellon University Language Technologies Institute Master of Science (M.S.) in Language Technologies > Advisors: Graham Neubig & Barnabás Póczos > Collaborators: Aditi Raghunathan , Ameet Talwalkar , David Mortensen > Honors: CMU LTI Research Fellowship	Pittsburgh, PA
May 2020 Aug 2016	Savitribai Phule Pune University PVG's College of Engineering and Technology Bachelor of Engineering (B.E.) in Computer Science	Pune, India

Experience

Aug 2026 May 2026	Microsoft Teams Core Applied Science <i>Research Intern Host: Henry Lai, Tanmay Gangwani & Eli Kravez</i> > Theoretical understanding of why adversarial training can fail for preference learning.	Redmond, WA
May 2026 Jan 2026	Thoughtworks <i>Student Researcher Host: Phillip Howard</i> > Representation geometry-based LLM post-training for enhancing generation diversity.	Remote
Aug 2024 May 2024	Apple Siri & Information Intelligence (SII) <i>Research Intern Host: Yuan Zhang & Hong Yu</i> <i>Collaborators: Joel Moniz, Xiou Ge, Bo-Hsiang Tseng, & Dhivya Piraviperumal</i> > Meta evaluation of hallucination detection metrics for LLM-generated texts (<i>Findings of EMNLP'25</i>).	Seattle, WA
Aug 2023 May 2023	Apple Siri & Information Intelligence (SII) <i>Research Intern Host: Shruti Bhargava & Hong Yu</i> <i>Collaborators: Bo-Hsiang Tseng, Joel Moniz, & Dhivya Piraviperumal</i> > Synthetic data generation using LLMs for dialog state tracking. (<i>EACL'24</i>)	Seattle, WA
July 2022 March 2021	IIIT Delhi Laboratory for Computational Social Systems (LCS2) [🌐] <i>Predoctoral Researcher Advisor: Tanmoy Chakraborty</i> > Hate speech detection under subpopulation and covariate shifts. (<i>KDD'23</i>) > Semantic role labeling of multimodal memes using optimal transport. (<i>EACL'23</i>) > Claim span identification using structural constraints and domain-specific rationales. (<i>EMNLP'22</i>) > Sarcasm explanation generation in multi-modal multi-party dialogues. (<i>ACL'22</i>)	Delhi, India

Selected Publications (* denotes equal contribution)

- [13] **Disentangling Geometry, Performance, and Training in Language Models** [PDF]
[Atharva Kulkarni](#), Jacob Springer, Arjun Subramonian, Swabha Swayamdipta
International Conference on Machine Learning (**Spotlight paper - top 2% of the submissions**) [ICML'26]
- [12] **Economies of Open Intelligence: Tracing Power & Participation in the Model Ecosystem** [PDF]
Shayne Longpre, Christopher Akiki, Campbell Lund, [Atharva Kulkarni](#), ..., Lucie-Aimée Kaffee
Transactions of Machine Learning Research | Media Coverage: [Financial Times](#) [TMLR'26]

- [11] **Evaluating Evaluation Metrics – The Mirage of Hallucination Detection** [PDF]
Atharva Kulkarni, Yuan Zhang, Joel Moniz, Xiou Ge, ..., Swabha Swayamdipta, Hong Yu
Findings of the Association for Computational Linguistics [Findings of EMNLP'25]
- [10] **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]
- [9] **Still Not Quite There! Evaluating Large Language Models for Comorbid Mental Health Diagnosis** [PDF | code]
Amey Hengle*, Atharva Kulkarni*, Shantanu Patankar, ..., Rashmi Gupta
Conference on Empirical Methods in Natural Language Processing [EMNLP'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 (Oral Presentation) [EACL'24]
- [7] **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 Linguistics [Findings of EMNLP'23]
- [6] **Counting the Bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model** [PDF | code]
Leonie Weissweiler, ..., Atharva Kulkarni, ..., Hinrich Schuetze, Kemal Oflazer, David Mortensen
Conference on Empirical Methods in Natural Language Processing [EMNLP'23]
- [5] **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 (AI for Social Good Award) [IJCAI'23]
- [4] **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]
- [3] **Characterizing Entities in Harmful Memes: Who is Hero, Villain, Victim?** [PDF | code]
Shivam Sharma*, Atharva Kulkarni*, ..., 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 (Oral Presentation) [ACL'22]

Academic Service

Conference Reviewer	NeurIPS 2026, ICML2026, COLM (2025-2026), ACL (2023-2025), EMNLP (2022-2025)
Workshop Reviewer	WiNLP 2024, NeurIPS SSLTheoryPractice 2023, NeurIPS RO-FoMo 2023

Talks

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

- > NiklasOPF AI podcast [video link] Nov 2024 – YouTube
- > EACL'24 Oral Presentation March 2024 – St. Julian's, Malta
- > Siri & Information Intelligence (SII), 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

Student Mentorship

Dhruv Tarsadiya	CS Masters, USC
Brennen Ho	CS Undergrad, USC

Teaching & Leadership

- Mentor – CS Undergraduate Mentorship Program | USC** Fall'24
‣ Mentoring USC undergraduate students for starting out in research (Mentee – Tshephang Makubate & Grace Li).
- Teaching Assistant – Advanced NLP | CMU** Fall'23
‣ TA for 11-711 [Advanced NLP](#) taught by [Prof. Daniel Fried](#). Designed assignments and guided teams for course projects.
- Teaching Assistant – Introduction to NLP | CMU** Spring'23
‣ TA for 11-611 [Intro to NLP](#) taught by [Prof. David Mortensen](#). Designed assignments and guided teams for course projects.
- Mentor – AI Undergraduate Mentorship Program | CMU** Fall'22
‣ Mentored CMU undergraduate students for starting out in research (Mentee – Yi (Nicole) Xiang).

Awards & Honours

1. **ICML 2026 Spotlight Paper** – Disentangling Geometry, Performance, and Training in Language Models
2. **ICML 2026 Silver Reviewer Award**
3. **USC Viterbi Fellowship (2024-25)** – Awarded to selected incoming PhD student at USC, covering full tuition & stipend.
4. **CMU LTI Research Fellowship (2023-24)** – Awarded to selected MLT students at CMU, covering full tuition & stipend.
5. **Best Paper – AI for Social Good | IJCAI'23** [🔗] – Poverty estimation using neural networks for rural Indian setting.
6. **Shared Task Winner | WASSA @ EACL'21** [🔗] – Empathy & distress detection in reaction to news stories.
5. **Shared Task Winner | TechDoFication @ ICON'20** [🔗] – Coarse-grained technical domain identification of documents for indigenous Indian languages.
6. **Bachelor's Thesis Award | ZS ASPIRE 2020** [🔗] – 2nd place amongst 33,000 teams & cash prize of 2,00,000 INR for Bachelor's Thesis project.

References

- | | |
|---------------------------------|---|
| Prof. Swabha Swayamdipta | <i>Assistant Professor, University of Southern California, USA</i> [🔗, @] |
| Prof. Barnabás Póczos | <i>Associate Professor, Carnegie Mellon University, USA</i> [🔗, @] |
| Prof. Graham Neubig | <i>Associate Professor, Carnegie Mellon University, USA</i> [🔗, @] |