

# Sunipa Dev

Mobile: +1 (214) 566-4559

Email: sunipadev@google.com

Website : sunipa.github.io

## AREAS OF SPECIALIZATION

---

*Language Modeling ★ Responsible AI ★ Machine Learning*

My research focuses on bridging gaps in AI models to ensure safe, and enriching human - AI interactions globally, irrespective of the identity of a person. I build and leverage interdisciplinary methods across my areas of expertise to build scalable, community-centered approaches for culturally aware AI.

## CURRENT POSITION

---

- **Google** Jan 2022 – Present  
*Senior Research Scientist, Google Research* *Seattle, WA*
  - Key role accomplishments: Lead projects and people to develop, and execute an end-to-end program for creation of culturally aware generative models. It involves (i) building community centered *and* scalable techniques for globally representative data, (ii) developing evaluation and mitigation strategies that embrace pluralism of perspectives to result in culturally responsive and meaningful generations across the world, and (iii) delivering these research grounded solutions to models and products across Google to improve human experiences with AI.

## EDUCATION AND TRAINING

---

- **University of California, Los Angeles** Jan 2021 – Jan 2022  
*NSF Computing Innovation Postdoctoral Fellow; PI : Kai-Wei Chang* *Los Angeles, CA*
- **University of Utah** Aug 2016 – Dec 2020  
*PhD Computer Science; Advisor : Jeff M Phillips* *Salt Lake City, UT*
- **Indian Institute of Science Education and Research, Kolkata** Jul 2011 – Jun 2016  
*Integrated Bachelors and Masters in Mathematics and Statistics* *Kolkata, India*

## PAST RESEARCH AND WORK EXPERIENCE

---

- **University of California, Los Angeles** Jan 2021 – Jan 2022  
*NSF Computing Innovation Postdoctoral Researcher; PI: Kai-Wei Chang* *Los Angeles, CA*
  - Proposed and executed novel research: Extend fundamental NLP tasks and applications to perform well beyond discrete identity categorizations such as binary gender. Understand and mitigate harmful model behaviors. Develop better evaluation paradigms for models towards socio-culturally aware NLP. Extend NLP tasks meaningfully to the healthcare domain and ensure fair mental health assessments across different socio-demographic groups.
- **University of Utah** Aug 2016 – Dec 2020  
*Research Assistant; Advisor: Jeff M Phillips* *Salt Lake City, UT*
  - Understand and enhance interpretability of high-dimensional representations. Analyze, evaluate, and mitigate social biases and incongruencies in language representations and their downstream applications in interpretable and cost-effective ways that do not break down embedding structures.
- **Microsoft** June 2020 – August 2020  
*Data Scientist Intern; Advisor: Yi Zhang* *Redmond, WA*
  - Built adaptive user representations which attend to a user's current needs for providing better Ad engagement.

- **Visa Research** May 2019 – Aug 2019  
*PhD Research Intern; Advisor: Wei Zhang* *Palo Alto, CA*
  - Built user and merchant representations in interpretable ways so as build better recommender systems to match users to merchants.
- **IISER Kolkata** June 2014 – June 2016  
*Senior Thesis Researcher; Advisor: Anirban Banerjee* *Kolkata, India*
- **Okinawa Institute of Science and technology** May 2015 – Aug 2015  
*Research Intern; Advisor: Robert Sinclair* *Okinawa, Japan*
- **Institute of Mathematical Sciences** May 2014 – Aug 2014  
*Research Assistant; Advisor: Rahul Siddhantan* *Chennai, India*

---

## SELECTED AWARDS

- Named 100 Brilliant Women in AI Ethics 2022 by the WAIE project.
- DAAD AI-net Postdoc Fellow 2021
- ACL Outstanding Paper Award 2023
- ACL Outstanding Reviewer 2021
- Computing Innovation Fellowship Award Recipient 2020 (by CRA, CCC and NSF)
- Grace Hopper Scholar 2019
- INSPIRE Fellowship awardee by DST and Govt. of India for top 1% students in Science. (2011-2016)

---

## SPEAKING ENGAGEMENTS

- Plenary Speaker on Multicultural AI at Inaugural COLM 2024
- Invited Panelist at Annual Meeting of American Mathematical Society 2025
- Keynote Speaker at Deployable AI Workshop at KDD 2024
- Invited Panelist at Trust and Safety in the Majority World 2024
- Invited Speaker at Women in Engineering Silicon Valley 2024
- Invited Talk at Language Technologies Institute at CMU 2024
- Invited Talk at NLP Workshop at KAIST 2024
- Keynote Speaker at IndoML 2023
- Keynote Speaker at TrustNLP Workshop at ACL 2023
- Invited Talk at UC Irvine 2022
- Invited Talk at University of Bocconi 2021
- Invited Talk at LinkedIn Research Seminar 2021

---

## SELECTED PUBLICATIONS

For the most updated and extended list of my publications, please refer to my Google Scholar page.

1. P. Lertvittayakumjorn, D. Kinney, V. Prabhakaran, D. Martin, **S. Dev** “Towards Geo-Culturally Grounded LLM Generations”; 2025
2. A. Davani\*, **S. Dev**\*, H. Urbina, V. Prabhakara; “A Comprehensive Framework to Operationalize Social Stereotypes for Responsible AI Evaluations”, 2025
3. M. Bhutani, K. Robinson, V. Prabhakaran, S. Dave, **S. Dev**; “SeeGULL Multilingual: a Dataset of Geo-Culturally Situated Stereotypes”; ACL 2024

4. A. Jha, V. Prabhakaran, R. Denton, S. Laszlo, S. Dave, R. Qadri, C. Reddy, **S. Dev**; “ViSAGE: A Global-Scale Analysis of Visual Stereotypes in Text-to-Image Generation”; ACL 2024
5. A. Davani, S. Gubbi, **S. Dev**, S. Dave, V. Prabhakaran; “GeniL: A Multilingual Dataset on Generalizing Language”; COLM 2024 (oral presentation)
6. K. Robinson, S. Kudugunta, R. Stella, **S. Dev**, J. Bastings; “MiTTenS: A Dataset for Evaluating Gender Mistranslation”; EMNLP 2024
7. **S. Dev**, J. Goyal, D. Tewari, S. Dave, V. Prabhakaran; “Building Socio-culturally Inclusive Stereotype Resources with Community Engagement”; NeurIPS Datasets and Benchmarks 2023
8. A. Jha, A. Davani, C. Reddy, S. Dave, V. Prabhakaran, **S. Dev**; “SeeGULL: A Stereotype Benchmark with Broad Geo-Cultural Coverage Leveraging Generative Models”; ACL 2023
9. N. Selvam, **S.Dev**, T. Knot, D. Khashabi, KW. Chang; “The Tail Wagging the Dog: Dataset Construction Biases of Social Bias Benchmarks”; ACL 2023; oral presentation and **Outstanding Paper Award**
10. A. Chowdhery, ..., **S. Dev**, ..., N. Fiedel; “PaLM: Scaling language modeling with pathways”; JMLR 2023
11. R. Anil, ..., **S.Dev**,..., Y. Wu; “Palm 2 Technical Report”; 2023
12. S.Bhatt, **S. Dev**, P. Talukdar, S. Dave, V. Prabhakaran; “Re-contextualizing Fairness in NLP: The Case of India”; AACL 2022, Taiwan
13. T.Hossain, **S. Dev\***, S. Singh\*; “MISGENDERED: Limits of Large Language Models in Understanding Pronouns”; ACL 2023 (\* joint last and advising authors)
14. T. Hossain,**S. Dev**, S. Singh; “MisgenderMender: A Community-Informed Approach to Interventions for Misgendering”; NAACL 2024
15. M. Diaz, **S. Dev**, E. Reif, E. Denton, V. Prabhakaran; “SoUnD Framework: Analyzing (So)cial Representation in (Un)structured (D)ata”; ACM AIES 2024
16. V. Gadiraju, S. Kane, **S. Dev**, A. Taylor, D. Wang, E. Denton, R. Brewer; ““I wouldn’t say offensive but...”: Disability-Centered Perspectives on Large Language Models”; FAccT 2023
17. S. Dev, A. Jha, J. Goyal, D. Tewari, S. Dave, V. Prabhakaran “Building Stereotype Repositories with Complementary Approaches for Scale and Depth”; C3NLP at EACL 2023
18. A. Ovalle, **S. Dev**, J. Zhao, M. Sarrafzadeh, KW. Chang; “Auditing Algorithmic Fairness in Machine Learning for Health with Severity-Based LOGAN”; AAAI 2023 7th International Workshop on Health Intelligence
19. **S.Dev**, A. Jha, J. Goyal, D. Tewari, S. Dave, V. Prabhakaran; “Building NLP Evaluation Resources with LLMs and Community Engagement”; Deep Learning Indaba 2023
20. **S. Dev**, E. Sheng, J. Zhao, A. Amstutz, J. Sun, Y. Hou, M. Sanseverino, J. Kim, A. Nishi, N. Peng, and KW. Chang; “On measures of Biases and Harms in NLP”; AACL 2022, Taiwan
21. V. Malik, **S. Dev**, A. Nishi, N. Peng, and KW. Chang; “Socially Aware Bias Measurements for Hindi Language Representations”; NAACL 2022, Seattle, USA
22. **S. Dev**, M. Monajatipoor, A. Ovalle, A. Subramonian, J.M. Phillips and KW. Chang; “Harms of Gender Exclusivity and Challenges in Non-Binary Representation in Language Technologies”; EMNLP 2021 (oral presentation)
23. **S. Dev**, T. Li, J.M. Phillips and V. Srikumar; “OSCAR: Orthogonal Subspace Correction and Rectification of Biases in Word Embeddings”; EMNLP 2021 (oral presentation)

24. A. Rathore, **S. Dev**, J.M. Phillips, V. Srikumar, Y. Zheng, CM. Yeh, J. Wang, W. Zhang, B. Wang; “VERB: Visualizing and Interpreting Bias Mitigation Techniques for Word Representations”; ACM TIIS 2024
25. D. Wu, W. Ahmed, **S. Dev**, KW. Chang; “Representation Learning for Resource-Constrained Keyphrase Generation”
26. J. Qu, LH. Li, J. Zhao, **S. Dev**, KW. Chang “DisinfoMeme: A Multimodal Dataset for Detecting Meme Intentionally Spreading Out Disinformation”
27. **S. Dev**; “The geometry of distributed representations for better alignment, attenuated bias, and improved interpretability”; PhD Dissertation; University of Utah 2020
28. **S. Dev**, T. Li, J.M. Phillips and V. Srikumar; “On Measuring and Mitigating Biased Inferences of Word Embeddings”; AAAI 2020, New York City, USA; (oral presentation)
29. **S. Dev**, S. Hassan, J.M. Phillips; “Closed Form Word Embedding Alignment”; ICDM 2019, Beijing, China; (oral presentation) **In Special Issue of KAIS for Best Papers from ICDM**
30. **S. Dev**, J.M. Phillips; “Attenuating Bias in Word Vectors”; AISTATS 2019, Okinawa, Japan; (oral presentation + student award)
31. **S. Dev** “Brain Data Analysis for Visual Processing and Categorization in the Human Cortex”; MS Thesis; IISER Kolkata 2016

---

## PATENTS

- Feature subspace isolation and disentanglement in merchant embeddings;**Sunipa Dev**, Yan Zheng, Wei Zhang; US Patent 12002052; June 2024
- System, Method, and Computer Program Product for Debiasing Embedding Vectors of Machine Learning Models;**Sunipa Dev**,Yan Zheng, Michael Yeh, Junpeng Wang, Wei Zhang, Archit Rathore; pending

---

## SERVICE

- Senior Area Chair NAACL 2025 and 2024, ACL ARR from 2024 to present
- Workshop Chair EMNLP 2025
- Ethics Chair EMNLP 2024
- Area Chair COLM 2024, COLING 2024, ACL 2023, NAACL 2023, EMNLP 2023, AACL 2023, ACL ARR since 2021 and more.
- *Founding organizer* for annually conducted **Cross Cultural Considerations in NLP** Workshop at \*CL (EACL 2023, ACL 2024, upcoming NAACL 2025).
- Affinity Workshop Chair NeurIPS 2022
- Workflow chair AAAI 2022
- Reviewer and ethics reviewer for AI conferences including AAAI, EACL, ACL, EMNLP, NeurIPS, ICML, ICLR, FAccT, etc., and Science, Nature, TMLR, and IEEE journals since 2019
- Organizer of workshops on Human Centered, Responsible And Globally Inclusive AI at KDD 2021, NAACL 2024, and ICLR 2024.
- Organizer for Widening NLP 2021 - 2023; Advisor 2024 - present.
- Mentor at LatinX in AI 2021.
- Graduate Student Advisory Committee Member at the School of Computing University of Utah from 2017 to 2018. Contributed to Retention, Promotion Tenure (RPT) procedures of faculty members.

## TEACHING

---

- Tutorial on Geo-Culturally Aware AI at LREC COLING 2024
- Tutorial on Contextualizing AI with Cross Cultural Perspectives at FAccT 2023
- CRAFT Session at FAccT 2022 on Collaboratively Developing Evaluation Frameworks for AI Harms
- Tutorial on Visualizing Bias Mitigation Approaches in High Dimensional Embeddings at AAAI 2021, KDD 2021
- Guest Lecture on Statistics of Building Cultural AI at University of Washington, Center for Statistics and Social Sciences 2025.
- Guest Lecture on Geo-Culturally Responsive NLP and AI Systems at USC Information Sciences Institute Fall 2023
- Guest Lecture on Responsible and Inclusive NLP at UCLA Spring 2021
- Lead Teaching Assistant for Data Mining (Spring 2018, 103 students): conducted lectures, tutorials, project evaluations, office hours and graded assignments.
- Lead Teaching Assistant for Foundations of Data Analysis (Fall 2017, 43 students): conducted lectures, tutorials, project evaluations, office hours and graded assignments.
- Teaching Assistant for Mathematics II (Spring 2016, 200 students): conducted tutorials, office hours and graded assignments and exam papers.

## RESEARCH ADVISING

---

I have had the opportunity to advise, collaborate with, and serve as a PhD Committee Member for the following students.

- **Laura Cabello Piqueras** Feb 2025 - present  
*Datalogisk Institut, Københavns Universitet (DIKU)* *PhD Thesis Committee Member*
- **Myra Cheng** Feb 2025 - present  
*PhD student, Stanford University* *Google Student Researcher*
- **Eddie Ungless** Oct 2024 - Jan 2025  
*PhD student, U Edinburgh* *Google Intern*
- **Tamanna Hossain** Aug 2022 - May 2024  
*PhD student, UC Irvine*
- **Akshita Jha** Aug - Dec 2022, May - Aug 2023  
*PhD student, Virginia Tech* *Google Student Researcher*
- **Nikil Selvam** July 2021 – Jan 2022  
*Undergraduate Student, UCLA*
- **Masoud Monajatipor** Jan 2021 – Jan 2022  
*PhD Student, UCLA*
- **Anaelia Ovalle** Feb 2021 – Jan 2022  
*PhD Student, UCLA*
- **Arjun Subramonian** Feb 2021 – Nov 2022  
*PhD Student, UCLA*
- **Vijit Malik** April 2021 – Jan 2022  
*Undergraduate Student, IIT Kanpur, India*