NIKITA SONI

https://www3.cs.stonybrook.edu/~nisoni/
Stony Brook, New York
https://scholar.google.com/citations?user=1w2rduoAAAAJ&hl=en

RESEARCH INTERESTS

Human Language Models, Human-Centered NLP, Natural Language Understanding, Human context in LLMs and NLP, Human-Centered LLMs, NLP x Psychology and Mental Health, Human and Social Factors, Temporal Human Context, Personalization.

EDUCATION

Stony Brook University, New York
Ph.D. in Computer Science (Major in Natural Language Processing) *Aug 2020 - 2025*
Advisor : **H. Andrew Schwartz** and **Niranjan Balasubramanian** GPA: 3.92/4.0
Master of Science in Computer Science (Thesis) *Aug 2019 -*
Advisor : **H. Andrew Schwartz** and **Niranjan Balasubramanian**

Krishna Engineering College, Uttar Pradesh Technical University, India *Jul 2008 -Jun 2012*
Bachelor of Technology in Computer Science GPA : 76.86%

RESEARCH INTERNSHIP EXPERIENCE

- **NLP Research Intern, Bloomberg AI Group** *Jul 2023- Oct 2023*
  Dynamic Conditional Text Generation (extending on the Human Language Modeling formulation)
- **Bocconi University, Milan, Italy** *Sep 2022 - Dec 2022*
  Visiting Researcher in Computing Science Department at MilaNLP Lab
  Advisor : **Dirk Hovy**
- **PhD NLP (Data Science) Intern, Capital One** *Jun 2021 - Aug 2021*
  Customer (dynamic) Sentiment Analysis.
- **Data Science Intern, McAfee LLC** *May 2020 - Aug 2020*
  Anomalous Sequences Detection in user activities for different cloud services.

PUBLICATIONS

21. **[Ongoing Work] Benchmarking Human-Context-aware Large Language Models**
Nikita Soni, Dhruv Vijay Kunjadiya, Pratham Piyush Shah, Dikshya Mohanty, H. Schwartz, Niranjan Balasubramanian

20. **[Ongoing Work; Dataset For SemEval-2026 Shared Task That I Am Leading] Longitudinal Dataset for Predicting Variation in Emotional Valence and Arousal over Time from Ecological Essays**
Nikita Soni, Tony Bui, Ryan Boyd, Syeda Mahwish, August Håkan Nilsson, Adithya V Ganesan, Lyle Ungar, Niranjan Balasubramanian1, Saif M. Mohammad, H. Andrew Schwartz

19. **Addressing the Ecological Fallacy in Larger LLMs with the Author's Context**
Nikita Soni, Dhruv Vijay Kunjadiya, Pratham Piyush Shah, Dikshya Mohanty, H. Schwartz, Niranjan Balasubramanian
In Submission (Long)

18. **[Technical Report To Be Submitted] LHLC: Large Human Language Data Corpus**
Nikita Soni, Dhruv Vijay Kunjadiya, Pratham Piyush Shah, Dikshya Mohanty, H. Schwartz, Niranjan Balasubramanian

17. **Residualized Similarity for Faithfully Explainable Authorship Verification**
Peter Zeng, Pegah Alipoormolabashi, Jihu Mun, Gourab Dey, **Nikita Soni**, Niranjan Balasubramanian, Owen Rambow, H. Andrew Schwartz
Findings of the Association for Computational Linguistics: EMNLP 2025 (Long)

16. **Systematic Evaluation of Auto-Encoding and Large Language Model Representations for Capturing Author States and Traits**
Khushboo Singh, Vasudha Varadarajan, Adithya V. Ganesan, August Håkan Nilsson, **Nikita Soni**, Syeda Mahwish, Pranav Chitale, Ryan L. Boyd, Lyle Ungar, Richard N. Rosenthal, H. Andrew Schwartz
Findings of the Association for Computational Linguistics: ACL 2025 (Long)

15. **Evaluation of LLMs-based Hidden States as Author Representations for Psychological Human-Centered NLP Tasks**
Nikita Soni, Pranav Chitale, Khushboo Singh Niranjan Balasubramanian, H. Andrew Schwartz
Findings of the Association for Computational Linguistics: NAACL 2025 (Short)

14. **Who We Are, Where We Are: Mental Health at the Intersection of Person, Situation, and Large Language Models**
Nikita Soni, August Håkan Nilsson, Syeda Mahwish, Vasudha Varadarajan, H. Andrew Schwartz, Ryan L. Boyd
Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2025) (NAACL 2025) (Short)

13. **The Consistent Lack of Variance of Psychological Factors Expressed by LLMs and Spambots**
Vasudha Varadarajan, Salvatore Giorgi, Siddharth Mangalik, **Nikita Soni**, Dave M. Markowitz, H. Andrew Schwartz
Proceedings of Detecting AI-Generated Content Workshop at COLING 2025 (Short)

12. **Large Human Language Models: A Need and the Challenges**
Nikita Soni, H. Andrew Schwartz, João Sedoc, Niranjan Balasubramanian
Proceedings of The 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Long)

11. **Comparing Pre-trained Human Language Models: Is it Better with Human Context as Groups, Individual Traits, or Both?**
Nikita Soni, Niranjan Balasubramanian, H. Andrew Schwartz, Dirk Hovy
Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis (WASSA 2024) (ACL 2024) (Long)

10. **Proceedings of the 1st Human-Centered Large Language Modeling Workshop**
Nikita Soni, Lucie Flek, Ashish Sharma, Diyi Yang, Sara Hooker, H. Andrew Schwartz
Proceedings of the 1st Human-Centered Large Language Modeling Workshop (ACL 2024)

9. **From Text to Context: Contextualizing Language with Humans, Groups, and Communities for Socially Aware NLP**
Adithya V Ganesan, Siddharth Mangalik, Vasudha Varadarajan, **Nikita Soni**, Swanie Juhng, João Sedoc, H. Andrew Schwartz, Salvatore Giorgi, Ryan L. Boyd
Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 5: Tutorial Abstracts)

8. **Automatic Implicit Motives Codings are at Least as Accurate as Humans' and 99% Faster**
   August Håkan Nilsson, J. Malte Runge, Oscar Kjell, **Nikita soni**, Adithya V Ganesan, and Carl Viggo Nilsson
   Journal of Personality and Social Psychology: Personality Processes and Individual Differences, 2024 (Journal)

7. **Robust language-based mental health assessments in time and space through social media**
   Siddharth Mangalik, Johannes C Eichstaedt, Salvatore Giorgi, Jihu Mun, Farhan Ahmed, Gilvir Gill, Adithya V Ganesan, Shashanka Subrahmanya, **Nikita Soni**, Sean AP Clouston, and H. Andrew Schwartz.
   NPJ Digital Medicine, 7(1):109, 2024. (Journal)

6. **Archetypes and Entropy: Theory-Driven extraction of Evidence for Suicide Risk**
   Vasudha Varadarajan, Allison Lahnala, Adithya V Ganesan, Gourab Dey, Siddharth Mangalik, Ana-Maria Bucur, **Nikita Soni**, Rajath Rao, Kevin Lanning, Isabella Valejo, Lucie Flek, H. Andrew Schwartz, Charles Welch, and Ryan L. Boyd.
   Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024) (EACL 2024) (Short)

5. **"I Slept Like a Baby": Using Human Traits To Characterize Deceptive ChatGPT and Human Text**
   Salvatore Giorgi, David M. Markovitz, **Nikita Soni**, Vasudha Varadarajan, Siddharth Mangalik, and H. Andrew Schwartz
   International workshop on implicit author characterization from texts for search and retrieval (IACT'23) (Long)

4. **Human Language Modeling**
   Nikita Soni, Matthew Matero, Niranjan Balasubramanian, and H. Andrew Schwartz
   Findings of the Association for Computational Linguistics: ACL 2022 (Long)

3. **WWBP-SQT-lite: Multi-level Models and Difference Embeddings for Moments of Change Identification in Mental Health Forums**
   Adithya V Ganesan, Vasudha Varadarajan, Juhi Mittal, Shashanka Subrahmanya, Matthew Matero, **Nikita Soni**, Sharath Chandra Guntuku, Johannes Eichstaedt, H. Andrew Schwartz
   Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2022) (NAACL 2022) (Long)

2. **Detecting Dissonant Stance in Social Media: The Role of Topic Exposure**
   Vasudha Varadarajan, **Nikita Soni**, Weixi Wang, Christian Luhmann, H. Andrew Schwartz, Naoya Inoue
   Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+ CSS 2022) (EMNLP 2022) (Short)

1. **MeLT: Message-Level Transformer with Masked Document Representations as Pre-Training for Stance Detection**
   Matthew Matero, **Nikita Soni**, Niranjan Balasubramanian, and H. Andrew Schwartz
   Findings of the Association for Computational Linguistics: EMNLP 2021 (Short)

PROFESSIONAL ACTIVITIES & SERVICE

- Organizing a **SemEval-2026 Shared Task: Predicting Variation in Emotional Valence and Arousal over Time from Ecological Essays**
  Nikita Soni, H. Andrew Schwartz, Tony Bui, Ryan Boyd, August Håkan Nilsson, Syeda Mahwish, Adithya V Ganesan, Lyle Ungar, Niranjan Balasubramanian, Saif M. Mohammad

- Organized the **1st Human-Centered Large Language Modeling Workshop** co-located with ACL 2024
  Nikita Soni, Lucie Flek, Ashish Sharma, Diyi Yang, Sara Hooker, and H. Andrew Schwartz.
- Tutorial on **From Text to Context: Contextualizing Language with Humans, Groups, and Communities for Socially Aware NLP** at NAACL 2024
  Adithya V Ganesan, Siddharth Mangalik, Vasudha Varadarajan, **Nikita Soni**, Swanie Juhng, João Sedoc, H. Andrew Schwartz, Salvatore Giorgi, and Ryan Boyd
- Organized a **Birds of Feather Session at ACL 2024** focused on **mentorship and community building for Human-Centered Large Language Modeling**.
- Organized a **Birds of Feather Session at ACL 2024** to engage in **an open dialogue on mental health challenges faced by graduate students, postdocs, faculty, and researchers**.
- Program Committee at EMNLP 2025, NAACL 2025, ACL 2024, NAACL 2024, NAACL 2023, EMNLP 2023, EMNLP 2022
- Program Committee at The 5th workshop on Natural Language Processing and Computational Social Science (NLP+CSS).
- Program Committee at ICWSM Data Challenge 2023
- Volunteer for Diversity & Inclusion Committee (LatinX IN AI) at NAACL 2022

STUDENT MENTORING

- Dhruv Vijay Kunjadiya: MS at SBU (2024 - present)
- Pratham Piyush Shah: MS at SBU (2024 - present)
- Dikshya Mohanty: PhD at SBU (2024 - present)
- Khushboo Singh: MS at SBU (2024-2025)
- Pranav Chitale: MS Thesis at SBU (2024-2025) → PhD student at Stony Brook University
- Shailen Smith (co-mentored): UG Thesis honors at SBU (2023-2024) → PhD student at Dartmouth College

TEACHING

- CSE 538: Graduate Natural Language Processing – Guest Lecture: Linear Regression. [Fall, 2020]
- CSE 357: UG Probability & Statistics for Data Science – Guest Lecture: Tensorflow. [Fall, 2021]
- CSE 538: Graduate Natural Language Processing – Teaching Assistant. [Fall, 2020]
- CSE 354: UG Natural Language Processing – Teaching Assistant. [Spring, 2020]

INVITED TALKS AND INTERVIEWS

- **Human-Centered Large Language Modeling**, STEM Speaker Series, University Libraries, Stony Brook University, Nov 2024
- **Large Human Language Models: A Need and the Challenges**, All Things Language and Computation, Stony Brook University, Oct 2024
- **CRA-WP Video Interview**, Interviewed by Computing Research Association - Widening Participation in Minnesota, April 2024 CRA-Women workshop
- **Transformers and Self-Attention**, Guest Lecture at H. Andrew Schwartz's Graduate NLP Class CSE 538, Stony Brook University, March 2024

- **Human Language Modeling**, Bloomberg AI Group, July 2023
- **Human Language Modeling**, MilaNLP Group, Bocconi University, Sep 2022
- **Human Language Modeling**, All Things Language and Computation, Stony Brook University, Apr 2022
- **Human Language Modeling**, World Well-Being Project (WWBP) Consortium, Mar 2022

PROFESSIONAL EXPERIENCE (INDIA)

- **Lead Software Development Engineer in Test, [24]7.ai** *Apr 2018 - Aug 2019*
- **Senior Software Development Engineer in Test, PegaSystems** *Mar 2015 - Apr 2018*
- **Software QA Engineer, McAfee Software** *Aug 2012 - Mar 2015*

REFERENCES

- **H. Andrew Schwartz**
  Associate Professor, College of Connected Computing
  Vanderbilt University, Nashville, TN, USA
  Email: has@cs.stonybrook.edu
- **Niranjan Balasubramanian**
  Associate Professor, Department of Computer Science
  Stony Brook University, New York, USA
  Email: niranjan@cs.stonybrook.edu
- **Dirk Hovy**
  Professor, Computing Sciences Department
  Bocconi University, Milan, Italy
  Email: dirk.hovy@unibocconi.it
- **Chris Callison-Burch**
  Professor, Computer and Information Science
  University of Pennsylvania, Pennsylvania, USA
  Email: ccb@upenn.edu
- **Owen Rambow**
  Professor, Department of Linguistics
  Stony Brook University, New York, USA
  Email: owen.rambow@.stonybrook.edu
- **João Sedoc**
  Assistant Professor of Technology, Operations, and Statistics
  New York University Stern School of Business, New York, USA
  Email: jsedoc@stern.nyu.edu