

Swabha Swayamdipta

POSTDOCTORAL INVESTIGATOR, ALLEN INSTITUTE FOR AI

✉ swabhas@allenai.org | 🏠 https://swabhs.com | 🌐 swabhs | 🐦 @swabhz

Research Interests

NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING

- Dataset Interpretability: Analyzing and improving data quality from the point of view of model training
- Model Interpretability: Analyzing predictive behavior of models
- Spurious and Social Biases in Datasets and Models
- Inductive Biases for Structured Prediction and Unsupervised Learning

Professional Experience

University of Southern California

GABILAN ASSISTANT PROFESSOR OF COMPUTER SCIENCE

Los Angeles, CA, USA

(Upcoming) Fall 2022

Allen Institute for AI

POSTDOCTORAL INVESTIGATOR

- **Advisor:** Yejin Choi
- **Project Team:** MOSAIC

Seattle, WA, USA

Aug 2019 - PRESENT

Allen Institute for AI

RESEARCH INTERN, **Host:** MATTHEW E. PETERS

- **Project:** Shallow Syntactic Priming of Large-scale Language Models

Seattle, WA, USA

Summer-Fall 2018

Google Inc.

RESEARCH INTERN, **Host:** ANKUR P. PARIKH

- **Project:** Machine Comprehension for Google Neon

New York, NY, USA

Summer 2017

Columbia University

RESEARCH ASSISTANT, CENTER FOR COMPUTATIONAL LEARNING SYSTEMS

New York, NY, USA

Jan 2013 - May 2013

Google Inc.

SOFTWARE ENGINEER INTERN

- Information retrieval evaluation for Google TV search

Mountain View, CA, USA

Summer 2012

Oracle Server Technologies

MEMBER TECHNICAL STAFF

- **Project:** Unified messaging for Fusion Middleware

Bangalore, India

2010 - 2011

Microsoft Corporation

SOFTWARE DEVELOPMENT INTERN

- **Project:** Search functionality for customer hierarchies

Hyderabad, India

Summer 2009

Education

Carnegie Mellon University

PHD IN LANGUAGE AND INFORMATION TECHNOLOGIES

- **Advisors:** Noah A. Smith, Chris Dyer
- **University of Washington**, Seattle, WA — Visiting PhD student Fall 2015 - Summer 2019
- **Thesis:** Syntactic Inductive Biases in NLP

Pittsburgh, PA, USA

2013 - 2019

Columbia University

MASTERS IN COMPUTER SCIENCE

- **Advisors:** Owen Rambow, Michael Collins
- **Specialization:** Natural Language Processing

City of New York, NY, USA

2011 - 2012

Teaching

TUTORIALS

2019	Transfer Learning in Natural Language Processing , Attendees: 350+	NAACL
2018	Frame Semantics across Languages: Towards a Multilingual FrameNet , Attendees: 30+	CoLING

GUEST LECTURES

Spring 2021	Transfer Learning , Host: Zaid Harachoui	<i>UW DATA 598: Statistical Deep Learning</i>
Fall 2020	Biases and Interpretability in NLP , Host: Eunsol Choi	<i>UT-Austin CS 395T: Topics in NLP</i>
Winter 2019	Phrase-Structure Parsing , Host: Noah A. Smith	<i>UW CSE 447/547M: NLP</i>
Spring 2018	Minimum Bayes Risk Decoding , Host: Waleed Ammar	<i>UW CSE 599 D1: Advanced Topics in NLP</i>
Spring 2018	Dependency Parsing , Host: Noah A. Smith	<i>UW CSEP 517: NLP</i>
Fall 2017	Unsupervised Learning , Host: Noah A. Smith	<i>UW CSE 446: Machine Learning</i>
Fall 2014	Dependency Parsing with Chu-Liu-Edmonds , Hosts: Chris Dyer, Alon Lavie, Bob Frederking	<i>CMU 11-271: Algorithms for NLP</i>

TEACHING ASSISTANTSHIP

Spring 2015	CMU 11-411/611: Natural Language Processing , Instructors: Chris Dyer, Alan W Black, Shomir Wilson	<i>Head TA</i>
-------------	---	----------------

Mentorship

Summer 2021-Present	Kawin Ethayarajh , PhD, Stanford University	<i>AI2 Intern</i>
Summer 2021-Present	Sarah Wiegrefe , PhD, George Institute of Technology	<i>AI2 Intern</i>
Summer 2021-Present	Ximing Liu , BS, UW CSE	<i>AI2 Intern</i>
Fall 2020-Present	Alisa Liu , PhD, UW CSE	
Fall 2020	Alon Jacovi , PhD, Bar Ilan University	<i>AI2 Intern</i>
Summer 2020-1	Jenny Liang , BS, UW CSE	<i>AI2 Intern</i>
Summer 2020	Jize Cao , BS, UW CSE	<i>AI2 Intern</i>
Spring 2020-1	Xuhui Zhou , CLMS, UW Linguistics	
Fall 2019 - Spring 2020	Yiben Yang , PhD, Northwestern University	
Fall 2019 - Spring 2020	Chaitanya Malaviya , PYI, Allen Institute for AI	
2018 - 2019	Ron Fan , MS, UW CSE	
2018 - 2019	Karishma Mandyam , BS, UW CSE	

Invited Talks

Nov 02, 2020	Responsible AI: Addressing Biases in Datasets and Models , E+D Product Leads	Microsoft
Oct 30, 2019	Biases and Learning Challenges in Natural Language Processing , Rising Stars in EECS	UIUC
Oct 18, 2019	Sprucing up a Dataset: Adversarially Filtering Dataset Artifacts , Linguistics Colloquium	UW Linguistics
Apr 08, 2019	Learning Challenges in Natural Language Processing , AI2 Seminar	Allen Institute for AI
Apr 04, 2019	Learning Challenges in Natural Language Processing , NLP Seminar	Stanford University
Mar 13, 2019	Learning Challenges in Natural Language Processing , CLSP Seminar	Johns Hopkins University
Mar 11, 2019	Learning Challenges in Natural Language Processing , CILVR Seminar	New York University
Jan 22, 2019	Learning Challenges in Natural Language Processing , MSR Seminar	Microsoft Research, New York
Sep 15, 2018	Representation Learning with Linguistic Structure , MSR AI Breakthroughs Workshop	Microsoft Research, Redmond
Apr 27, 2018	Syntactic Scaffolds for Semantic Structures , NorthWest NLP	Microsoft Research, Redmond

Awards and Honors

2020	Best Paper: Honorable Mention , Don't Stop Pretraining: Adapt Language Models to Domains and Tasks	ACL
2009	Sun Campus Ambassador Scholarship , Bangalore, India	Sun Microsystems
2003	National Talent Search Scholarship , Rourkela, India	NCERT, India

Publications

CONFERENCE PAPERS

Contrastive Explanations for Model Interpretability A. Jacovi, S. Swayamdipta , S. Ravfogel, Y. Elazar, Y. Choi, and Y. Goldberg	EMNLP 2021
On-the-Fly Controlled Text Generation with Experts and Anti-Experts A. Liu, M. Sap, X. Lu, S. Swayamdipta , C. Bhagavatula, N. A. Smith, and Y. Choi	ACL 2021
Challenges in Automated Debiasing for Toxic Language Detection • X. Zhou, M. Sap, S. Swayamdipta , N. A. Smith, and Y. Choi	EACL 2021
Dataset Cartography: Mapping and Diagnosing Datasets with Training Dynamics • S. Swayamdipta , R. Schwartz, N. Lourie, Y. Wang, H. Hajishirzi, N. A. Smith, Y. Choi	EMNLP 2020
Generative Data Augmentation for Commonsense Reasoning • Y. Yang, C. Malaviya, J. Fernandez, S. Swayamdipta , R. LeBras, J. Wang, C. Bhagavatula, Y. Choi, and D. Downey	Findings of EMNLP 2020
Adversarial Filters of Dataset Biases • R. LeBras, S. Swayamdipta , C. Bhagavatula, R. Zellers, M. E. Peters, A. Sabharwal, and Y. Choi	ICML 2020
The Right Tool for the Job: Matching Model and Instance Complexities • R. Schwartz, G. Stanovsky, S. Swayamdipta , J. Dodge, and N. A. Smith	ACL 2020
Don't Stop Pretraining: Adapt Language Models to Domains and Tasks • S. Gururangan, A. Marasović, S. Swayamdipta , K. Lo, I. Beltagy, D. Downey, and N. A. Smith • Best Paper Honorable Mention	ACL 2020
Syntactic Scaffolds for Semantic Structures • S. Swayamdipta , S. Thomson, K. Lee, L. Zettlemoyer, C. Dyer, N. A. Smith.	EMNLP 2018
Learning Joint Semantic Parsers from Disjoint Data • H. Peng, S. Thomson, S. Swayamdipta , N. A. Smith	NAACL 2018
Annotation Artifacts in Natural Language Inference Data • S. Gururangan*, S. Swayamdipta *, O. Levy, R. Schwartz, S. Bowman, and N. A. Smith • *equal contribution	NAACL 2018
Polyglot Semantic Role Labeling • P. Mulcaire, S. Swayamdipta , and N. A. Smith	ACL 2018
Multi-Mention Learning for Reading Comprehension with Neural Cascades • S. Swayamdipta , A. Parikh, T. Kwaitkowski	ICLR 2018
Greedy, Joint Syntactic and Semantic Parsing with Stack LSTMs • S. Swayamdipta , M. Ballesteros, C. Dyer, N. A. Smith	CoNLL 2016
A Dependency Parser for Tweets • L. Kong, N. Schneider, S. Swayamdipta , A. Bhatia, C. Dyer, N. A. Smith	EMNLP 2014
The Pursuit of Power and its Manifestation in Written Dialog • S. Swayamdipta , O. Rambow	ICSC 2012

WORKSHOP PAPERS

Sister Help: Data Augmentation for Frame-Semantic Role Labeling • A. Pancholy, S. Swayamdipta , M. R. L. Petruck	LAW-DMR @ EMNLP 2021
Multi-Task Learning for Incremental Parsing using Stack LSTMs • S. Swayamdipta , M. Ballesteros, C. Dyer, N. A. Smith	WiML @ NeurIPS 2016

- CMU: Arc-Factored, Discriminative Semantic Dependency Parsing SemEval 2014
- S. Thomson, D. Bamman, J. Dodge, **S. Swayamdipta**, N. Schneider, C. Dyer, N. A. Smith
- The CMU Machine Translation Systems WMT 2014
- A. Matthews, C. Dyer, A. Lavie, G. Hanneman, W. Ammar, A. Bhatia, **S. Swayamdipta**, E. Schlinger, Y. Tsvetkov

WORKING PAPERS

- MAUVE: Human-Machine Divergence Curves for Evaluating Open-Ended Text Generation arXiv:2102.01454
- K. Pillutla, **S. Swayamdipta**, R. Zellers, J. Thickstun, Y. Choi, and Z. Harchaoui
- Shallow Syntax in Deep Water arXiv:1908.11047
- **S. Swayamdipta**, M. Peters, B. Roof, C. Dyer and N. A. Smith
- Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold arXiv:1706.09528
- **S. Swayamdipta**, S. Thomson, C. Dyer and N. A. Smith
- DyNet: The Dynamic Neural Network Toolkit arXiv:1701.03980
- G. Neubig, C. Dyer, Y. Goldberg, A. Matthews, W. Ammar, A. Anastasopoulos, M. Ballesteros, D. Chiang, D. Clothiaux, T. Cohn, K. Duh, M. Faruqi, C. Gan, D. Garrette, Y. Ji, L. Kong, A. Kuncoro, G. Kumar, C. Malaviya, P. Michel, Y. Oda, M. Richardson, N. Saphra, **S. Swayamdipta**, P. Yin

TUTORIAL AND ORGANIZATION PAPERS

- Transfer Learning in Natural Language Processing NAACL 2019
- S. Ruder, M. E. Peters, **S. Swayamdipta**, T. Wolf
- Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019) EMNLP 2019
- C. Cherry, G. Durrett, G. Foster, R. Haffari, S. Khadivi, N. Peng, X. Ren, **S. Swayamdipta**
- Frame Semantics across Languages: Towards a Multilingual FrameNet CoLING 2018
- C. F. Baker, M. Ellsworth, M. R. L. Petruck, **S. Swayamdipta**

Professional Service

Area Chair

- EMNLP 2021: Machine Learning for NLP
- NAACL 2021: Sentence-Level Semantics
- EACL 2021: Sentence-Level Semantics
- ACL 2020: Semantics (Long)

Reviewer: Conferences

- | | |
|-----------|-----------------|
| • ACL | 2015-2020 |
| • NAACL | 2015-2020 |
| • EMNLP | 2015-2020 |
| • NeurIPS | 2018-2020 |
| • ICML | 2015,2019,2020 |
| • EACL | 2017 |
| • AAAI | 2017-2020 |
| • CoNLL | 2017-2018, 2020 |

Reviewer: Workshops

- | | |
|--|------|
| • NAACL Student Research Workshop | 2016 |
| • Workshop for Women in Machine Learning (WiML) at NeurIPS | 2016 |
| • Machine Reading for Question Answering (MRQA) Workshop | 2017 |

Reviewer: Journals

- | | |
|-----------------------------|----------------|
| • Computational Linguistics | 2019 - PRESENT |
| • Transactions of ACL | 2020 - PRESENT |
| • Journal of AI Research | 2019 |

Co-organizer

- | | |
|--|-----------|
| • EMNLP 2019: Workshop on Deep Learning in Low Resource NLP (DeepLo) | 2019 |
| • West Coast NLP Workshop | 2018 |
| • Women's Research Day | 2017 |
| • UW-NLP Retreat | 2015-2016 |

Mentorship at Conferences

- EMNLP 2020, ACL 2020, EMNLP 2018

COMMUNITY SERVICE

Allen Institute for AI

- Diversity, Equity and Inclusion Committee
- AllenAI Outstanding Engineer Scholarship Committee

Seattle WA, USA

2019-PRESENT
2019

Paul G. Allen School of CSE, University of Washington

- Faculty Recruiting Liaison
- Graduate Student Advisory Council
- Outreach: Highline Public High School
- CS: Gradswomen

Seattle WA, USA

2019
2018
2017
2016-2019

References

Yejin Choi, yejin@cs.washington.edu

UW; AI2

Noah A. Smith, noah@cs.washington.edu

UW; AI2

Chris Dyer, cdyer@google.com

Google DeepMind

Luke Zettlemoyer, lsz@cs.washington.edu

UW; FAIR