Robin Jia

Email: robinjia@usc.edu Address: 941 Bloom Walk, Los Angeles, CA 90089
Website: https://robinjia.github.io/ Office: SAL 236 Phone: (213) 740-5906

Education

Stanford University September 2014 – August 2020
Ph.D. in Computer Science
Advisor: Percy Liang
Thesis: Building Robust Natural Language Processing Systems

Stanford University September 2010 – June 2014
Bachelor of Science with Honors in Computer Science
Minor in Biology
GPA: 4.103 / 4.0

Employment

Assistant Professor, Computer Science Department, University of Southern California
Los Angeles, CA August 2021 – Present

Visiting Researcher, Facebook AI Research
Menlo Park, CA August 2020 – August 2021
Hosts: Luke Zettlemoyer and Douwe Kiela

Research Intern, Microsoft Research
Redmond, WA June 2018 – September 2018
Host: Hoifung Poon

Research Intern, Google Research
Mountain View, CA June 2016 – September 2016
Host: Larry Heck

Awards

Google Research Scholar Award	2023 – 2024
Cisco Research Award	2023 – 2024
Open Philanthropy research grant	2021 – 2024
Best Short Paper	ACL 2018
Outstanding Paper (Best paper runner-up)	EMNLP 2017
National Science Foundation Graduate Research Fellow	2014 – 2019

Publications

Do Localization Methods Actually Localize Memorized Data in LLMs?
Ting-Yun Chang, Jesse Thomason, and Robin Jia NAACL 2024

Efficient End-to-End Visual Document Understanding with Rationale Distillation
Wang Zhu, Alekh Agarwal, Mandar Joshi, Robin Jia, Jesse Thomason, and Kristina Toutanova NAACL 2024

Chain-of-Questions Training with Latent Answers for Robust Multistep Question Answering
Wang Zhu, Jesse Thomason, and Robin Jia EMNLP 2023

SCENE: Self-Labeled Counterfactuals for Extrapolating to Negative Examples
Deqing Fu, Ameya Godbole, and Robin Jia EMNLP 2023

Estimating Large Language Model Capabilities without Labeled Test Data
Harvey Yiyun Fu, Qinyuan Ye, Albert Xu, Xiang Ren, and Robin Jia EMNLP Findings 2023

How Predictable Are Large Language Model Capabilities? A Case Study on BIG-bench
Qinyuan Ye, Harvey Yiyun Fu, Xiang Ren, and Robin Jia EMNLP Findings 2023

Data Curation Alone Can Stabilize In-context Learning
Ting-Yun Chang and Robin Jia ACL 2023

Contrastive Novelty-Augmented Learning: Anticipating Outliers with Large Language Models
Albert Xu, Xiang Ren, and Robin Jia
ACL 2023
SoCalNLP Symposium 2022 Best Paper Award.

Are Sample-Efficient NLP Models More Robust?
Nelson F. Liu, Ananya Kumar, Percy Liang, and Robin Jia
ACL 2023

Do Question Answering Modeling Improvements Hold Across Benchmarks?
Nelson F. Liu, Tony Lee, Robin Jia, and Percy Liang
ACL 2023

Does VLN Pretraining Work with Nonsensical or Irrelevant Instructions?
Wang Zhu, Ishika Singh, Yuan Huang, Robin Jia, and Jesse Thomason
O-DRUM 2023

Benchmarking Long-tail Generalization with Likelihood Splits
Ameya Godbole and Robin Jia
EACL Findings 2023

Generalization Differences between End-to-End and Neuro-Symbolic Vision-Language Reasoning Systems
Wang Zhu, Jesse Thomason, and Robin Jia
EMNLP Findings 2022

Knowledge base question answering by case-based reasoning over subgraphs
Rajarshi Das, Ameya Godbole, Ankita Naik, Elliot Tower, Manzil Zaheer, Hannaneh Hajishirzi, Robin Jia, and Andrew McCallum
ICML 2022

On the Robustness of Reading Comprehension Models to Entity Renaming
Jun Yan, Yang Xiao, Sagnik Mukherjee, Bill Yuchen Lin, Robin Jia, and Xiang Ren
NAACL 2022

Models in the Loop: Aiding Crowdworkers with Generative Annotation Assistants
Max Bartolo, Tristan Thrush, Sebastian Riedel, Pontus Stenetorp, Robin Jia, and Douwe Kiela
NAACL 2022

Question Answering Infused Pre-training of General-Purpose Contextualized Representations
Robin Jia, Mike Lewis, and Luke Zettlemoyer
ACL Findings 2022

Analyzing Dynamic Adversarial Training Data in the Limit
Eric Wallace, Adina Williams, Robin Jia, and Douwe Kiela
ACL Findings 2022

On Continual Model Refinement in Out-of-Distribution Data Streams
Bill Yuchen Lin, Sida Wang, Xi Victoria Lin, Robin Jia, Lin Xiao, Xiang Ren, and Scott Yih
ACL 2022

Dynaboard: An Evaluation-As-A-Service Platform for Holistic Next-Generation Benchmarking
Zhiyi Ma*, Kawin Ethayarajh*, Tristan Thrush*, Somya Jain, Ledell Wu, Robin Jia, Christopher Potts, Adina Williams, and Douwe Kiela
NeurIPS 2021

Masked Language Modeling and the Distributional Hypothesis: Order Word Matters Pre-training for Little
Koustuv Sinha, Robin Jia, Dieuwke Hupkes, Joelle Pineau, Adina Williams, and Douwe Kiela
EMNLP 2021

Improving Question Answering Model Robustness with Synthetic Adversarial Data Generation
Max Bartolo, Tristan Thrush, Robin Jia, Sebastian Riedel, Pontus Stenetorp, and Douwe Kiela
EMNLP 2021

To What Extent do Human Explanations of Model Behavior Align with

Actual Model Behavior?
Grusha Prasad, Yixin Nie, Mohit Bansal, Robin Jia, Douwe Kiela, and Adina Williams
BlackBoxNLP 2021

The statistical advantage of automatic NLG metrics at the system level
Johnny Tian-Zheng Wei and Robin Jia
ACL 2021

Evaluation Examples Are Not Equally Informative:
How Should That Change NLP Leaderboards?
Pedro Rodriguez, Joe Barrow, Alexander Hoyle, John P. Lalor, Robin Jia, and Jordan Boyd-Graber
ACL 2021

Do Explanations Help Users Detect Errors in Open-Domain QA?
An Evaluation of Spoken vs. Visual Explanations
Ana Valeria Gonzalez, Gagan Bansal, Angela Fan, Yashar Mehdad, Robin Jia, and Srinivasan Iyer
ACL Findings 2021

Swords: A Benchmark for Lexical Substitution with
Improved Data Coverage and Quality
Mina Lee*, Chris Donahue*, Robin Jia, Alexander Iyabor, and Percy Liang
NAACL 2021

Dynabench: Rethinking Benchmarking in NLP
Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia, Zhiyi Ma, Tristan Thrush, Sebastian Riedel, Zeerak Waseem, Pontus Stenetorp, Robin Jia, Mohit Bansal, Christopher Potts, and Adina Williams
NAACL 2021

On the Importance of Adaptive Data Collection for
Extremely Imbalanced Pairwise Tasks
Stephen Mussmann*, Robin Jia*, and Percy Liang
EMNLP Findings 2020

With Little Power Comes Great Responsibility
Dallas Card, Peter Henderson, Urvashi Khandelwal, Robin Jia, Kyle Mahowald, and Dan Jurafsky
EMNLP 2020

Selective Question Answering under Domain Shift
Amita Kamath, Robin Jia, and Percy Liang
ACL 2020

Robust Encodings: A Framework for Combating Adversarial Typos
Erik Jones, Robin Jia*, Aditi Raghunathan*, and Percy Liang
ACL 2020

Certified Robustness to Adversarial Word Substitutions
Robin Jia, Aditi Raghunathan, Kerem Göksel, Percy Liang
EMNLP 2019

MRQA 2019 Shared Task: Evaluating Generalization in Reading Comprehension
Adam Fisch, Alon Talmor, Robin Jia, Minjoon Seo, Eunsol Choi, and Danqi Chen
MRQA 2019

Document-Level N-ary Relation Extraction with
Multiscale Representation Learning
Robin Jia, Cliff Wong, and Hoifung Poon
NAACL 2019

Know What You Don't Know: Unanswerable Questions for SQuAD
Pranav Rajpurkar*, Robin Jia*, and Percy Liang
ACL 2018
Best Short Paper Award.

Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer
Juncen Li, Robin Jia, He He, and Percy Liang
NAACL 2018

Adversarial Examples for Evaluating Reading Comprehension Systems
Robin Jia and Percy Liang
EMNLP 2017

Outstanding Paper Award.

Learning Concepts through Conversations in Spoken Dialogue Systems
Robin Jia, Larry Heck, Dilek Hakkani-Tür, and Georgi Nikolov ICASSP 2017

Data Recombination for Neural Semantic Parsing
Robin Jia and Percy Liang ACL 2016

”Reverse Genomics” Predicts Function of Human Conserved Noncoding Elements
Amir Marcovitz, Robin Jia, and Gill Bejerano MBE 2016

Mx1 and Mx2 Key Antiviral Proteins are Surprisingly Lost in Toothed Whales
Benjamin A. Braun, Amir Marcovitz, J. Gray Camp, Robin Jia, and Gill Bejerano PNAS 2015

Note: * denotes equal contribution.

Grants

Google Research Scholar Award, Google, $60,000
PI: Robin Jia May 2023 – May 2024
Title: *Stabilizing In-Context Learning by Understanding the Value of Demonstrations*

Cisco Research Award, Cisco, $70,000
PI: Robin Jia Apr 2023 – Apr 2024
Title: *Estimating Large Language Model Capabilities without Labeled Data*

Open Philanthropy Research Grant, Open Philanthropy, $320,000
PI: Robin Jia Aug 2021 – Aug 2024
Title: *Adversarial Robustness Research*

Students Supervised

Ph.D. students

Johnny Tian-Zheng Wei	Jan 2021 – present	
Ameya Godbole	Aug 2021 – present	
Wang (Bill) Zhu	Oct 2021 – present	Joint with Jesse Thomason
Ting-Yun (Charlotte) Chang	Jan 2022 – present	Joint with Jesse Thomason
Deqing Fu	Mar 2023 – present	Joint with Vatsal Sharan

Masters and undergraduate students

Daniel Firebanks-Quevedo (USC MS)	Jan 2024 – Present	
Gustavo Adolpho Lucas De Carvalho (USC MS)	Jan 2024 – Present	
Tianyu Wang (USC UG)	Aug 2023 – Dec 2023	
Rahel Selemon (Brown UG)	Jun 2023 – Aug 2023	
Qilin Ye (USC UG)	Jun 2023 – Present	
Ryan Wang (USC UG)	Apr 2023 – Present	Provost’s Research Fellowship
Tianqi Chen (USC MS)	Mar 2023 – Present	
Anthony Sauer (USC UG)	Jan 2023 – Mar 2024	
Yuan Huang (USC MS)	Jun 2022 – Oct 2023	
Harvey Fu (USC UG)	May 2022 – Present	Provost’s Research Fellowship
Adam Reynolds (USC MS)	Aug 2021 – Dec 2021	
Amita Kamath (Stanford MS)	Sep 2018 – Jun 2020	Now: UCLA Ph.D. student
Erik Jones (Stanford UG)	Jun 2019 – Dec 2019	Now: UC Berkeley Ph.D. student
Kerem Göksel (Stanford MS)	Jan 2019 – Jun 2019	Now: Semantic Machines

Faculty research project mentor for CAIS++ students Leslie Moreno, Aryan Gulati, and Aditya Kumar.

Ph.D. qualifying exam committee member: Hexiang Hu, Yury Zemlyanskiy, Zalan Fabian, Negar Mokhberian, Michiel de Jong, Pei Zhou, Qinyuan Ye, Jun Yan, Ming-Chang Chiu, Xisen

Jin, Fei Wang, Yun Cheng Wang, Soumya Sanyal, Jake Bremerman, Johnny Wei, Bingyi Zhang, Ali Omrani, Lee Kezar, Wang Zhu, Brihi Joshi, Ting-Yun Chang, Justin Cho (22 total).

Ph.D. thesis proposal committee member: Hexiang Hu, Yury Zemlyanskiy, Yuchen Lin, Aaron Chan, Wenxuan Zhou, Michiel de Jong, Woojeong Jin, Ming-Chang Chiu, Jun Yan, Fei Wang, Xisen Jin, Ali Omrani (12 total).

Ph.D. thesis defense committee member: Yury Zemlyanskiy, Hanqing Zeng, Aaron Chan, Wenxuan Zhou, Yun-Cheng Wang (5 total).

M.S. thesis committee member: Jeong Hyun An, Abid Hassan (2 total).

Teaching

Instructor, CSCI 467: Introduction to Machine Learning
University of Southern California, Los Angeles, CA Spring 2023, Fall 2023, and Spring 2024

Instructor, CSCI 699: Generalization and Robustness in Natural Language Processing
University of Southern California, Los Angeles, CA Spring 2022

Teaching Fellow, CS 221 (Artificial Intelligence)
Stanford University, Stanford, CA Summer 2019
Instructor for 100-student class on artificial intelligence.

Teaching Assistant, CS 124 (Introduction to Natural Language Processing)
Stanford University, Stanford, CA Winter 2018

Head Teaching Assistant, CS 221 (Artificial Intelligence)
Stanford University, Stanford, CA Fall 2015
Head TA of 550-student class, managed a team of 18 TA's.

Section Leader, CS 106A (Introduction to Programming)
Stanford University, Stanford, CA Winter 2012
Taught a section of twelve students, graded assignments and exams.

Tutor, Stanford University Mathematical Organization
Stanford University, Stanford, CA Winter 2011 – Spring 2012
Tutored students in linear algebra and multivariable calculus. Served as coordinator of the tutoring program from September 2011 to June 2012.

Professional Service

General Chair for 2023 SoCal NLP Symposium.

Steering Committee member for Workshop on Instruction Tuning and Instruction Following at NeurIPS 2023.

Co-instructor of Tutorial on Uncertainty Estimation for Natural Language Processing at COLING 2022.

Co-organizer of the First Workshop on Dynamic Adversarial Data Collection (DADC) at NAACL 2022.

Co-organizer of the Third Workshop on Machine Reading and Question Answering (MRQA) at EMNLP 2021.

Co-instructor of Tutorial on Robustness and Adversarial Examples in Natural Language Processing at EMNLP 2021.

Co-organizer of the Second Workshop on Machine Reading and Question Answering (MRQA) at EMNLP 2019.

Co-organizer of the First Workshop on Machine Reading and Question Answering (MRQA) at

ACL 2018.

Area chair for ACL (2021, 2023, 2024), EMNLP (2021, 2022, 2023), NAACL (2021), and AKBC (2021).

Reviewer for ACL Rolling Review (2021, 2022, 2023, 2024), ACL (2018, 2019, 2020), EMNLP (2018, 2019, 2020), NAACL (2019), TACL (2022, 2023, 2024), EACL (2022), AACL (2020), ICML (2019), CoNLL (2018), AKBC (2019, 2022), RobustSeq Workshop (2022), ML Safety Workshop (2022), DistShift Workshop (2021, 2022, 2023), BlackboxNLP Workshop (2021, 2022, 2023), Repl4NLP Workshop (2021, 2023), GenBench Workshop (2023), ACL Student Research Workshop (2021), RobustML Workshop (2021), EMNLP DeepLo Workshop (2019), and NAACL GenDeep Workshop (2018). Outstanding Reviewer for EMNLP 2020.

NSF panel reviewer.

Invited Talks

Knowing Machines Podcast	Oct 2023
CHAI Workshop Plenary Talk	Jun 2023
Capital One Research Invited Talk	Jun 2023
UC Irvine AI/ML Seminar	May 2022
Amazon Research Invited Talk	Apr 2022
Princeton NLP Group Seminar	Jul 2021
NLP Highlights Podcast	Mar 2021
USC ISI Seminar	Feb 2021
UC Santa Barbara NLP Seminar	Feb 2021

Other

Frederick Emmons Terman Engineering Scholastic Award, Stanford University	2014
Finalist, Lunsford Oral Presentation of Research Award, Stanford University	2012
Finalist, Boothe Prize for Excellence in Writing, Stanford University	2011
Top 500 Scorer, William Lowell Putnam Mathematical Competition	2011
Three-time Qualifier, USA Mathematics Olympiad	2008–2010
Top Twenty Finalist, US National Chemistry Olympiad	2009