Saadia K. Gabriel

Homepage: https://homes.cs.washington.edu/~skgabrie
Email: skgabrie@cs.washington.edu
GitHub: https://github.com/skgabriel

Education

University of Washington Fall 2017 – Current
PhD in Computer Science & Engineering
Dissertation Advisers: Prof. Yejin Choi and Prof. Franziska Roesner

University of Washington December 2019
Master's in Computer Science & Engineering

Mount Holyoke College, Summa Cum Laude May 2017
BA in Computer Science & Mathematics
Thesis Adviser: Prof. Dan Sheldon

Selected Honors and Achievements:

- Awarded Google-LEAP Fellowship (2021)
- Best Paper at WeCNLP summit as co-author of Social Bias Frames (2020)
- IROS Best Paper Nomination as co-author of Early Fusion for Goal Directed Robotic Vision (2019)
- ACL Best Paper Nomination as co-author of The Risk of Racial Bias in Hate Speech Detection (2019)
- CRA-W Grad Cohort Workshop Participant (2018)
- CRA URMD Grad Cohort Workshop Participant (2018)
- Phi Beta Kappa (2017)
- Weaver Award for Computer Science and Math (2017)
- David Notkin Endowed Graduate Fellowship in Computer Science & Engineering (2017)
- ARCS Foundation Fellowship (2017)

Publications and Presentations:

Peer-reviewed:

- **NaturalAdversaries: Can Naturalistic Adversaries Be as Effective as Artificial Adversaries?**

Saadia Gabriel, Hamid Palangi, Yejin Choi.
EMNLP 2022 Findings.

- **Misinfo Reaction Frames: Reasoning about Readers' Reactions to News Headlines**

Saadia Gabriel, Skyler Hallinan, Maarten Sap, Pemi Nguyen, Franziska Roesner, Eunsol Choi, Yejin Choi.
ACL 2022.

- **ToxiGen: A Large-Scale Machine-Generated Dataset for Adversarial and Implicit Hate Speech Detection**

Thomas Hartvigsen, **Saadia Gabriel**, Hamid Palangi, Maarten Sap, Dipankar Ray, Ece Kamar.
ACL 2022.

- **GO FIGURE: A Meta Evaluation of Factuality in Summarization**

Saadia Gabriel, Asli Celikyilmaz, Rahul Jha, Yejin Choi, Jianfeng Gao.
ACL 2021 Findings.

- **Discourse Understanding and Factual Consistency in Abstractive Summarization**

Saadia Gabriel, Antoine Bosselut, Jeff Da, Ari Holtzman, Jan Buys, Kyle Lo, Asli Celikyilmaz, Yejin Choi.
EACL 2021.

- **Paragraph-level Commonsense Transformers with Recurrent Memory**

Saadia Gabriel, Chandra Bhagavatula, Vered Shwartz, Ronan Le Bras, Maxwell Forbes, Yejin Choi.
AAAI 2021.

- **Social Bias Frames: Reasoning about Social and Power Implications of Language**

Maarten Sap, **Saadia Gabriel**, Lianhui Qin, Dan Jurafsky, Noah A. Smith, Yejin Choi.
ACL 2020.

- **Detecting and Tracking Communal Bird Roosts in Weather Radar Data**

Zezhou Cheng, **Saadia Gabriel**, Pankaj Bhambhani, Daniel Sheldon, Subhransu Maji, Andrew Laughlin, David Winkler.
AAAI 2020.

- **The Risk of Racial Bias in Hate Speech Detection**

Maarten Sap, Dallas Card, **Saadia Gabriel**, Yejin Choi, Noah A. Smith.
ACL 2019.

- **MathQA: Towards Interpretable Math Word Problem Solving with Operation-Based Formalisms**

Aida Amini, **Saadia Gabriel**, Shanchuan Lin, Rik Koncel-Kedziorski, Yejin Choi, Hannaneh Hajishirzi.
NAACL 2019.

- **Early Fusion for Goal Directed Robotic Vision**

Aaron Walsman, Yonatan Bisk, **Saadia Gabriel**, Dipendra Misra, Yoav Artzi, Yejin Choi, Dieter Fox.
IROS 2019.

In Submission:

- **Can Machines Learn Morality? The Delphi Experiment**

Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Jon Borchardt, **Saadia Gabriel**, Yulia Tsvetkov, Oren Etzioni, Maarten Sap, Regina Rini, Yejin Choi.
Arxiv 2022.

- **Commonsense Dialogic Question Generation**

Pedro Colon-Hernandez, **Saadia Gabriel**, Yejin Choi, Cynthia Breazeal and Hae Won Park.

Experience:

Microsoft NLP Research Intern (Summer 2021)

- Worked on detection of malicious code with Jay Stokes and hate speech detection with Hamid Palangi, Dipankar Ray and Ece Kamar

Microsoft Deep Learning Research Intern (Summer 2020)

- Worked on evaluation of factual consistency in generation with Asli Celikyilmaz and Rahul Jha

Ai2 Research Intern (Fall 2019 – Summer 2020)

- Worked on extracting and integrating commonsense knowledge as a member of the Mosaic team led by Yejin Choi

Computer Vision & Learning Intern, SRI International (Summer 2019)

- Worked with Ajay Divakaran and Karan Sikka on using pre-trained models for commonsense knowledge extraction and integrating commonsense knowledge into multimodal applications of NLP, including visual question answering and generation

Graduate Research Assistant, University of Washington (Fall 2017 – Current)

- Researching machine learning techniques and implementing deep-learning models for natural language understanding, social commonsense and logical reasoning in text
- Investigating ways of representing effects of actions in stories dependent on logical reasoning, like math word problems

Data Science Research Assistant, University of Massachusetts Amherst (Summer 2016)

- Developed computer vision models to identify bird roosts in radar data
- Worked with SQL and JavaScript to display results of roost detection in web application
- Participated in UMass College of Information and Computer Sciences poster presentation

GEM CS Mentor (Google-Funded Program), Mount Holyoke College (Spring 2016)

- Developed an active learning plan for Intro to Object-Oriented Programming class
- Mentored students and gave feedback in CS 101 lab
- Reviewed students' code and gave feedback on assignments

REU Research Assistant, University of Massachusetts Amherst (Summer 2015)

- Analyzed large datasets using Python and Matlab
- Developed parametric model to identify birds in radar data
- Presented research to technical and non-technical audiences

Wearable Electronics Inventor (2013 – Current)

- Created a jacket called The Turtle that charges mobile devices
- Gave a presentation on The Turtle and wearable technology for Computer Science Week at Mount Holyoke in Fall 2013

CS Educator (2012 – Current)

- Developed interactive movie application for teaching intro computer science and discrete mathematics to beginning students using hand-drawn animation

Teaching:

- TA for Real Analysis (Math 301), Mount Holyoke College
- TA for Undergrad NLP (CS 447) and Graduate NLP (CSE 517), University of Washington

Service:

- Reviewer for NeurIPS 2022
- Reviewer for EMNLP 2022
- Generation Session Chair for NAACL 2022
- Socio-Cultural Inclusion Co-Chair for NAACL 2022
- ARR Reviewer, Outstanding Reviewer at NAACL 2022
- Reviewer for Artificial Intelligence journal
- Reviewer for Computational Linguistics journal
- Reviewer for AAAI
- Reviewer for ACL 2020 (Outstanding Reviewer)
- Secondary Reviewer (ICLR 2019 and EMNLP 2019)
- PC for EACL SRW 2021, ACL SRW 2019 and NAACL 2019 workshops (NeuralGen and WNU)
- UW NLP Retreat Organizer (2018, 2019)
- UW CSE Visit Days Committee (2018)
- Mount Holyoke College CS Department Chair Student Search Committee (2016 - 2017)

Selected Talks:

- Invited talk at Cornell
  Modeling misinformation and implied toxicity to build less biased systems – April 2022
- Invited talk at Stanford
  Modeling misinformation, implied toxicity and commonsense implications with machine reasoning - December 2021

- MIT Rising Stars presentation
  Misinfo Reaction Frames – October 2021

- Invited talk at UMass Amherst
  Social Commonsense for Social Good - January 2021

- NeurIPS 2020 Resistance AI Workshop
  Social and Power Implications of Language - December 2020

- Colloquium Talk at University of Washington
  Social and Power Implications of Language – October 2020

- Invited Talk at BBN Technologies
  Cooperative Generator-Discriminator Networks – September 2020

- Voice Tech Global Panelist
  Implicit Bias in Conversational AI – July 2020

- Invited Talk at Mount Holyoke College
  MathQA to Co-opNet: Can We Teach NLP Models to Reason? – November 2019

- Invited Talk at Carlson School of Management
  NLP State-of-the-Art Methods – November 2019

- UW Quals Talk
  Co-opNet: Cooperative Generator-Discriminator Networks – October 2019

- MSR 2019 PhD Summit Poster Session
  Universal Frameworks for Commonsense Knowledge Integration – October 2019

- MSR 2018 PhD Summit Poster Session
  Neural Detox – October 2018

- Undergrad Thesis Defense
  Modeling Swallow Roosts Using Weather Radar – May 2017

- Mount Holyoke Computer Science Week
  The Turtle: A Solar-powered Jacket for Charging Mobile Devices – December 2013

Skills:

Programming: Python, Java, Matlab, R, ActionScript, HTML, C, JavaScript, SQL
Language: English (Native Speaker), French (Intermediate)