Alexis J. Ross

alexisro@mit.edu https://alexisjihyeross.github.io/

EDUCATION

Massachusetts Institute of Technology, Cambridge, Massachusetts

Sept 2022 -

Ph.D. in EECS

Harvard University, Cambridge, Massachusetts

Sept 2016 - May 2020

A.B. with Honors in Computer Science and Philosophy (Joint)

RESEARCH EXPERIENCE Allen Institute for AI (AI2), AllenNLP

July 2020 - July 2022

Predoctoral Researcher

Mentors: Ana Marasović and Matthew E. Peters

Harvard University, Computer Science and Philosophy

Sept 2019 - May 2020

Mentors: Himabindu Lakkaraju and Bernhard Nickel

Senior Thesis: Using Linear Approximations to Explain Complex, Blackbox Classifiers

Microsoft Research May 2019 – Aug 2019

Research Intern, Project Hanover

Mentor: Tristan Naumann

Project: Machine Reading for Precision Medicine

Harvard University, NLP Group

March 2019 - May 2019

Research Assistant

Mentors: Yonatan Belinkov and Alexander Rush

Project: Probing for Neurons that Encode and Exploit Negation in BERT

Johns Hopkins, Center for Language and Speech Processing

June 2018 – Aug 2018

Undergraduate Researcher, JSALT Workshop

Mentor: Ellie Pavlick

Project: General-Purpose Sentence Representation Learning

Harvard University, Cognitive Neuropsychology Laboratory

Feb 2017 - Aug 2017

Research Assistant

Mentors: Gilles Vannuscorps and Alfonso Caramazza

Project: Object Orientation Perception

PUBLICATIONS

[1] Does Self-Rationalization Improve Robustness to Spurious Correlations?. **Alexis Ross**, Matthew Peters, Matt Gardner, Ana Marasović *EMNLP*, 2022.

[2] Tailor: Generating and Perturbing Text with Semantic Controls.

Alexis Ross*, Tongshuang Wu*, Hao Peng, Matthew Peters, Matt Gardner. ACL, 2022.

[3] Learning Models for Actionable Recourse.

Alexis Ross, Himabindu Lakkaraju, and Osbert Bastani. NeurIPS, 2021.

[4] Competency Problems: On Finding and Removing Artifacts in Language Data.

Matt Gardner*, William Merrill*, Jesse Dodge, Matthew E. Peters, **Alexis Ross**, Sameer Singh, Noah Smith. *EMNLP*, 2021.

 $[5]\,$ Explaining NLP Models via Minimal Contrastive Editing (MiCE).

Alexis Ross, Ana Marasović, and Matthew E. Peters. Findings of ACL, 2021.

[6] How well do NLI models capture verb veridicality?

Alexis Ross and Ellie Pavlick. EMNLP 2019. Oral presentation.

[7] Probing what different NLP tasks teach machines about function word comprehension.

Najoung Kim, Roma Patel, Adam Poliak, Alex Wang, Patrick Xia, R Thomas McCoy, Ian Tenney,

 ${\bf Alexis}\ {\bf Ross},\ {\bf Tal}\ {\bf Linzen},\ {\bf Benjamin}\ {\bf Van}\ {\bf Durme},\ {\bf Samuel}\ {\bf R}\ {\bf Bowman},\ {\bf and}\ {\bf Ellie}\ {\bf Pavlick}.$ *SEM 2019. Best Paper Award.

SELECTED HONORS	MIT Shillman Fellowship Hoopes Prize "for extraordinary undergraduate research," awarded for senior thesis EMNLP Student Travel Scholarship Certificate of Distinction in Teaching for a student evaluation of 4.8/5.0
SERVICE	 Mentor, MIT EECS Graduate Application Assistance Program (GAAP) Co-organizer and frequent host, NLP Highlights Podcast, over 9000 views in October Co-organizer and frequent host, NLP Highlights Podcast, over 9000 views in October Co-organizer, AllenNLP Hacks, a hackathon for minoritized students Presented workshop: "My First NLP Project" Secondary Reviewer, ACL Mentor, Harvard Women in Computer Science
TEACHING EXPERIENCE	CS 121: Introduction to the Theory of Computation Teaching Fellow, Harvard School of Engineering and Applied Sciences Instructor: Boaz Barak Sept 2018 – Dec 2018
Other Experience	Solo, Chamber, and Orchestral Violinist 2002 – 2020 2016 – 2018 Violinist in the Harvard-Radcliffe Orchestra 2016 – 2020 Chamber Music Performance at Harvard. Coaches: Parker Quartet 2014 – 2016 Juilliard Pre-College, Diploma in Violin Performance. Teacher: Ann Setzer 2013 – 2015 Summers at the Meadowmount School of Music
	Mental Health Advocacy and Counseling, Harvard University 2017 – 2020 Events Co-Chair, Student Mental Health Liaisons 2018 – 2020 Peer Mental Health Counselor, Room 13