

# Alexis J. Ross

alexisr@allenai.org  
<https://alexisjihyeross.github.io/>

---

EDUCATION	<b>Harvard University</b> , Cambridge, Massachusetts A.B. with Honors in Computer Science and Philosophy (Joint)	Sept 2016 – May 2020
RESEARCH EXPERIENCE	<b>Allen Institute for Artificial Intelligence (AI2), AllenNLP</b> Predoctoral Researcher <i>Mentors:</i> Ana Marasović and Matthew E. Peters	July 2020 – Present
	<b>Harvard University, Computer Science and Philosophy</b> <i>Mentors:</i> Himabindu Lakkaraju and Bernhard Nickel <i>Senior Thesis:</i> Explaining Blackbox Models with Linear Approximations Awarded Thomas T. Hoopes Prize for “outstanding scholarly work” by undergraduates.	Sept 2019 – May 2020
	<b>Microsoft Research</b> Research Intern, Project Hanover <i>Mentor:</i> Tristan Naumann <i>Project:</i> Machine Reading for Precision Medicine	May 2019 – Aug 2019
	<b>Harvard University, NLP Group</b> Research Assistant <i>Mentors:</i> Yonatan Belinkov and Alexander Rush <i>Project:</i> Probing for Neurons that Encode and Exploit Negation in BERT	March 2019 – May 2019
	<b>Johns Hopkins, Center for Language and Speech Processing</b> Undergraduate Researcher, JSALT Workshop <i>Mentor:</i> Ellie Pavlick <i>Project:</i> General-Purpose Sentence Representation Learning	June 2018 – Aug 2018
	<b>Harvard University, Cognitive Neuropsychology Laboratory</b> Research Assistant <i>Mentors:</i> Gilles Vannuscorps and Alfonso Caramazza <i>Project:</i> Object Orientation Perception	Feb 2017 – Aug 2017
PREPRINTS	<ul style="list-style-type: none"><li>[1] Tailor: Generating and Perturbing Text with Semantic Controls. <b>Alexis Ross*</b>, Tongshuang Wu*, Hao Peng, Matthew Peters, Matt Gardner. In submission, 2021.</li><li>[2] Competency Problems: On Finding and Removing Artifacts in Language Data. Matt Gardner*, William Merrill*, Jesse Dodge, Matthew E. Peters, <b>Alexis Ross</b>, Sameer Singh, Noah Smith. In submission, 2021.</li><li>[3] Ensuring Actionable Recourse via Adversarial Training. <b>Alexis Ross</b>, Himabindu Lakkaraju, and Osbert Bastani. In submission, 2021.</li></ul>	
PUBLICATIONS	<ul style="list-style-type: none"><li>[4] Explaining NLP Models via Minimal Contrastive Editing (MiCE). <b>Alexis Ross</b>, Ana Marasović, and Matthew E. Peters. <i>Findings of ACL 2021</i>.</li><li>[5] How well do NLI models capture verb veridicality? <b>Alexis Ross</b> and Ellie Pavlick. <i>EMNLP 2019</i>. <b>Oral presentation</b>.</li><li>[6] 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, <b>Alexis Ross</b>, Tal Linzen, Benjamin Van Durme, Samuel R Bowman, and Ellie Pavlick. <i>*SEM 2019</i>. <b>Best Paper Award</b>.</li></ul>	

TEACHING  
EXPERIENCE

**CS 121: Introduction to the Theory of Computation**

Sept 2018 – Dec 2018

Teaching Fellow, Harvard School of Engineering and Applied Sciences

*Instructor:* Boaz Barak

Awarded a Certificate of Distinction in Teaching for a student evaluation of 4.8/5.0.