SHAULI RAVFOGEL

(+972) 0549508315 ⋄ shauli.ravfogel@gmail.com ⋄ shauli-ravfogel.netlify.app

Third-year Ph.D student at BIU NLP lab, supervised by Prof. Yoav Goldberg. My research interests encompass representation learning, unsupervised machine learning, causal attribution, and interpretability of neural models for NLP.

EDUCATION

Ph.D in Natural Language Processing and Machine Learning   2020 – Present
Bar-Ilan University
Advisor: Prof. Yoav Goldberg
Research Topics: Controlled Representation Learning; Bias Mitigation; Interpretability; Causality

M.Sc in Natural Language Processing and Machine Learning   2018 – 2020
Bar-Ilan University
Advisor: Prof. Yoav Goldberg
Research Topics: Syntax Acquisition in Neural Networks; Typological Diversity

B.Sc in Computer Science   2015 – 2018
Bar-Ilan University
Graduated with a final grade of 90.
Final Project: Implicit Syntax Learning in Morphologically-Rich Languages.

B.Sc in Chemistry   2010 – 2013
Bar-Ilan University
Graduated with a final grade of 90; Dean's Honors in 2011.

EXPERIENCE

Machine Learning Lecturer   Oct 2021 - March 2021

Teaching an introduction to machine learning course for brain science students.

Ph.D. Research Intern, AI2 Israel   Jan 2020 - Present

Working as a research intern on a project which focuses on combining symbolic and neural-based IE.

Machine Learning Teaching Assistant   May 2019 - 2021

TA in a graduate-level machine learning course for brain science students.

Medicinal Chemist   Sep 2013 - Sep 2016

Research position focused on civilian defense and protection against chemical and biological threats.

PUBLICATIONS

- Royi Rassin*, **Shauli Ravfogel***, Yoav Goldberg. **DALLE-2 is Seeing Double: Flaws in Word-to-Entity Mapping**. BlackBoxNLP@EMNLP 2022.

- **Shauli Ravfogel**, Francisco Vargas, Yoav Goldberg and Ryan Cotterell. **Adversarial Concept Erasure in Kernel Space**. EMNLP 2022.

- Hila Gonen, **Shauli Ravfogel**, and Yoav Goldberg. **Analyzing Gender Representation in Multilingual Models.** Proceedings of the 7th Workshop on Representation Learning for NLP 2022 (Best Paper Award).

- **Shauli Ravfogel**, Michael Twiton, Yoav Goldberg and Ryan Cotterell. **Linear Adversarial Concept Erasure.** International Conference on Machine Learning. PMLR, 2022.

- Elad Ben-Zaken, **Shauli Ravfogel**, and Yoav Goldberg. **Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models**. ACL 2022.

- **Shauli Ravfogel***, Grusha Prasad*, Tal Linzen and Yoav Goldberg. **Counterfactual Interventions Reveal the Causal Effect of Relative Clause Representations on Agreement Prediction**. CoNLL 2021.

- Alon Jacovi, Swabha Swayamdipta, **Shauli Ravfogel**, Yanai Elazar, Yejin Choi, Yoav Goldberg, **Contrastive Explanations for Model Interpretability**. EMNLP 2021.

- **Shauli Ravfogel**, Hillel Taub-Tabib, Yoav Goldberg. **Neural Extractive Search**. ACL 2021 demo track.

- **Shauli Ravfogel**, Yanai Elazar, Jacob Goldberger and Yoav Goldberg. **Unsupervised Distillation of Syntactic Information from Contextualized Word Representations**. BlackBoxNLP@EMNLP 2020.

- Carlo Meloni*, **Shauli Ravfogel*** , and Yoav Goldberg. **Ab Antiquo: Neural Proto-language Reconstruction**. NAACL 2021.

- Yanai Elazar, **Shauli Ravfogel**, Alon Jacovi and Yoav Goldberg. **Amnesic Probing: Behavioral Explanation with Amnesic Counterfactuals**. Accepted in TACL 2021.

- Hila Gonen, **Shauli Ravfogel**, Yanai Elazar and Yoav Goldberg. **It's not Greek to mBERT: Inducing Word-Level Translations from Multilingual BERT**. BlackBoxNLP 2020.

- **Shauli Ravfogel**, Yanai Elazar, Hila Gonen, Micheal Twiton and Yoav Goldberg. **Null It Out: Guarding Protected Attributes by Iterative Nullspace Projection**. ACL 2020.

- **Shauli Ravfogel**, Yoav Goldberg and Tal Linzen. **Studying the Inductive Biases of RNNs with Synthetic Variations of Natural Languages**. NAACL-HLT 2019.

- **Shauli Ravfogel**, Francis M. Tyers and Yoav Goldberg. **Can LSTM Learn to Capture Agreement? The Case of Basque**. BlackboxNLP@EMNLP 2018.

INVITED TALKS

- October 2021, the SIGTYP lecture series. Linear Controlled Representation Learning.

- December 2020, prof. Roi Reichart's Research Group, Technion. Assessing the Causal Effect of Linguistic Concepts with Counterfactual Representations.

- March 2020, Yale Linguistics Department. Unsupervised Distillation of Syntactic Information from Neural Representations.

- February 2020, Johns Hopkins Computation and Psycholinguistics Lab. Neutralizing Concepts in the Representations of Neural Models.

- February 2019, ONLP lab, the Open University of Israel. Implicit Syntax Acquisition by NNs in Morphologically Rich Languages.

ACADEMIC SERVICE

Reviewer: ACL 2020, CoNLL 2020, HAMLETS@NeurIPS2020, NAACL 2021, ACL-IJCNLP 2021, EMNLP 2021, ARR 2021-2022, ICML 2022

AWARDS

- 2022-2023: Bloomberg Data Science PhD Fellowship.

- 2022: Israeli Council for Higher Education (Vatat) Fellowship for outstanding PhD Students in Data Science (declined).

- 2022: Best paper award, RepL4NLP workshop.

- 2021-2024: The Presidential Scholarship for Outstanding PhD Students.

- 2020-2021: The Departmental Scholarship for Outstanding Doctoral Fellows.

- 2011: Dean's Honors (Chemistry)