

ORION WELLER

Baltimore Maryland, 21218

<http://orionweller.github.io> ♦ wellerorion@gmail.com

EDUCATION

Johns Hopkins University

Ph.D. in Computer Science

Center for Language and Speech Processing

Advisors: Benjamin Van Durme and Dawn Lawrie

2021 - Present

GPA: 4.0/4.0

Brigham Young University

Bachelor of Science, Summa Cum Laude (Top 1%)

Majors: Computer Science, Statistics; Minor in Mathematics

2017 - 2021

GPA: 4.0/4.0

AWARDS

- NSF Graduate Research Fellowship, 2022
- DoD NDSEG Fellowship (declined), 2022
- CRA Outstanding Undergraduate Researcher in North America Award, 2021
- Fulbright Fellowship Semi-Finalist, 2021
- NSF Graduate Research Fellowship Honorable Mention, 2021
- Goldwater Scholarship, 2020
- BYU Scholarships: Wessel/Marshall Memorial (2019), Juanita Miller Nelson (2018)
- BYU Dean's List (all semesters)

PREPRINTS

- [1] **Orion Weller**, Marc Marone, Nathaniel Weir, Dawn Lawrie, Daniel Khashabi, Benjamin Van Durme, ““According to...” Prompting Language Models Improves Quoting from Pre-Training Data,” *ArXiv*, 2023.
- [2] **Orion Weller**, Dawn Lawrie, Benjamin Van Durme, “NevIR: Negation in Neural Information Retrieval,” *ArXiv*, 2023.
- [3] James Mayfield, Eugene Yang, Dawn Lawrie, Samuel Barham, **Orion Weller**, Marc Mason, Suraj Nair, Scott Miller, “Synthetic Cross-language Information Retrieval Training Data,” *ArXiv*, 2023.
- [4] Samuel Barham, **Orion Weller**, Michelle Yuan, Kenton Murray, Mahsa Yarmohammadi, Zhengping Jiang, Siddharth Vashishtha, Alexander Martin, Anqi Liu, Aaron Steven White, “MegaWika: Millions of reports and their sources across 50 diverse languages,” *ArXiv*, 2023.
- [5] **Orion Weller**, Aleem Khan, Nathaniel Weir, Dawn J Lawrie, Benjamin Van Durme, “Defending Against Poisoning Attacks in Open-Domain Question Answering,” *ArXiv*, 2022.
- [6] Kangda Wei, Dawn J Lawrie, Benjamin Van Durme, Yunmo Chen, **Orion Weller**, “When Do Decompositions Help for Machine Reading?” *ArXiv*, 2022.

PUBLICATIONS

- [7] **Orion Weller**, Marc Marone, Vladimir Braverman, Dawn Lawrie, Benjamin Van Durme, “Pre-trained Models for Multilingual Federated Learning,” in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022.
- [8] **Orion Weller**, Kevin Seppi, Matt Gardner, “When to Use Multi-Task Learning vs Intermediate Fine-Tuning for Pre-Trained Encoder Transfer Learning,” in *Association of Computational Linguistics (ACL)*, 2022.
- [9] **Orion Weller**, Matthias Sperber, Telmo Pires, Hendra Setiawan, Christian Gollan, Dominic Telaar, Matthias Paulik, “Speech Translation for Code-Switched Speech,” in *Findings of the Association of Computational Linguistics (ACL)*, 2022.
- [10] Wilson Fearn, **Orion Weller**, Kevin Seppi, “Exploring the Relationship Between Algorithm Performance, Vocabulary, and Run-Time in Text Classification,” in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.
- [11] **Orion Weller**, Matthias Sperber, Christian Gollan, Joris Kluivers, “Streaming Joint Speech Translation and Transcription,” in *European Chapter of the Association for Computational Linguistics (EACL)*, Apr. 2021.
- [12] **Orion Weller**, Nicholas Lourie, Matt Gardner, Matthew Peters, “Learning from Task Descriptions,” in *Empirical Methods in Natural Language Processing (EMNLP)*, Nov. 2020.
- [13] **Orion Weller**, Jordan Hildebrandt, Ilya Reznik, Christopher Challis, E. Shannon Tass, Quinn Snell, Kevin Seppi, “You Don’t Have Time to Read This: an Exploration of Document Level Reading Time Prediction,” in *Association of Computational Linguistics (ACL)*, Jul. 2020.
- [14] **Orion Weller**, Nancy Fulda, Kevin Seppi, “Can Humor Prediction Datasets be used for Humor Generation? Humorous Headline Generation via Style Transfer,” in *Second Workshop on Figurative Language Processing @ ACL*, Jul. 2020.
- [15] **Orion Weller**, Luke Sagers, Carl Hanson, Quinn Snell, Michael Barnes, Shannon Tass, “Predicting Suicidal Thoughts and Behavior among Adolescents using the Risk and Protective Factor Framework: a Large-Scale Machine Learning Approach,” in *Plos One*, Nov. 2021.
- [16] **Orion Weller**, Kevin Seppi, “The rJokes Dataset: a Large Scale Humor Collection,” in *Language Resources and Evaluation (LREC)*, Mar. 2020.
- [17] **Orion Weller**, Kevin Seppi, “Humor Detection: a Transformer Gets the Last Laugh,” in *Empirical Methods in Natural Language Processing (EMNLP)*, Nov. 2019.

RESEARCH EXPERIENCE

- | | |
|--|---|
| Graduate Research Assistant at JHU | Sept 2021 - Present |
| <ul style="list-style-type: none">• Studying various information seeking NLP tasks [5] [6]. Work published at NAACL’21 [7]. | |
| Apple AI/ML Research | June 2020 - Sept 2020; May 2021 - Sept 2021 |
| <ul style="list-style-type: none">• Developed efficient and multilingual speech translation models. Our work has been presented at WeCNLP’20 and published at EACL’21 [11] and Findings of ACL’22 cs-st | |
| Allen Institute for Artificial Intelligence | Jan 2020 - June 2020 |
| <ul style="list-style-type: none">• Developed a framework towards general language understanding, by learning to solve tasks from their descriptions. Work published at ENNLP’20 [12] and ACL’22 [8]. | |
| Applied Machine Learning Lab at BYU | Dec 2018 - Dec 2019; Sept 2020 - May 2021 |

- Researched computational humor, gathering a dataset of more than half a million jokes. Developed models to identify and understand humor, showing improved results via transfer learning. Work published at EMNLP'19 [17], LREC'20 [16] and FigLang@ACL workshop [14].

Computational Health Lab at BYU

Sept 2018 - Dec 2019

- Studied psychological predictors of suicidal ideation. Work presented at the Society of Prevention Research 2020 Conference, to the Utah Prevention Department, and published in Plos One [15].

Undergraduate Capstone Project with Adobe

Sept 2018 - Apr 2019

- Researched NLP modeling of humans (psycholinguistics) to analyze how humans read document-sized text. Work published at ACL'20 [13].

ENGINEERING EXPERIENCE

Qualtrics

Apr 2019 - Aug 2019

- Responsible for prototyping a new system for handling edits on the Data Pipeline team.

Digi International

Oct 2017 - Jan 2019

- Worked on building firmware for wireless networking modules on the Zigbee standard.

SKILLS

Languages

English (native), Portuguese (intermediate), French (basic)

Programming Languages

Python, C++, R, Javascript, Golang, Java, OCaml, SQL

Frameworks & Tools

PyTorch, Linux, Mechanical Turk, Django, VueJS

SERVICE

Program Committee Member

- 2023: ACL, EMNLP, ACL Rolling Review, MASC-SLL (JHU organizer)
- 2022: ACL Rolling Review, MASC-SLL (JHU organizer)
- 2021: ACL Rolling Review, ACL, EMNLP, NAACL (secondary)
- 2020: Meta-Learning@NeurIPS

After School Coding Instructor for Elementary School

2021 - 2022

Develop curriculum and teach an after school class to 15 3rd-5th graders in Baltimore City with the Code in the Schools program.