

Sarah Wiegrefe

Atlanta, GA

Email: swiegrefe6@gatech.edu

Website: sarahwie.github.io

EDUCATION

Georgia Institute of Technology (Georgia Tech) August 2017 - 2022

PhD in Computer Science

Advisor: Prof. Mark Riedl

Passed qualifying exams: Spring 2019

M.S. in Computer Science (Machine Learning specialization) May 2020

Honors College at the College of Charleston August 2013 - May 2017

Bachelor of Science in Data Science, Summa Cum Laude

Minors in Mathematics and International Studies

University of Tartu, Estonia January - June 2015

Visiting Student, Faculty of Mathematics and Computer Science

PUBLICATIONS (acceptance rates listed where known)

Wiegrefe, S., Marasović, A. and Smith, Noah A. *Measuring Association Between Labels and Free-Text Rationales*. In submission.

Jain, S., **Wiegrefe, S.**, Pinter, Y. and Wallace, B. *Learning to Faithfully Rationalize by Construction*. Long Paper (all oral presentations). ACL 2020. Acceptance rate: 22.7%.

Wiegrefe, S. and Pinter, Y. (equal contribution). *Attention is not not Explanation*. EMNLP 2019. Long paper. Oral Presentation. Acceptance rate: 24% (7% oral presentations).

Wiegrefe, S., Flores, G., Choi, E., and Dai, A. *Learning Bi-Directional Clinical Event Representations: a Comparison of Architectures*. Preprint. 2019.

Wiegrefe, S., Choi, E., Yan, S., Sun, J. and Eisenstein, J. *Clinical Concept Extraction for Document-Level Coding*. ACL BioNLP Workshop 2019. Long paper.

Mullenbach, J., **Wiegrefe, S.**, Duke, J., Sun, J. and Eisenstein, J. *Explainable Prediction of Medical Codes from Clinical Text*. NAACL 2018. Long paper. Oral Presentation. Acceptance rate: 31%.

Wiegrefe, S., Anderson, P. and Obeid, J. *Can Classifications of Publications by Translational Categories be Automated?*. American Medical Informatics Association (AMIA) Joint Summits on Translational Science 2017.

RESEARCH

Research Intern, Allen Institute for Artificial Intelligence**May - October 2020**

Worked with Ana Marasović and Noah A. Smith on the AllenNLP team on quantifying faithful rationalization from free-text neural models.

Research Intern, Google AI Health (formerly Medical Brain)**May - August 2019**

Worked with Dr. Edward Choi and Gerardo Flores on the Medical Records team to improve on outcome prediction for clinical time-series using unsupervised pretraining techniques.

Computational Linguistics Lab at Georgia Tech**August 2017 - July 2019**

Advised by Professor Jacob Eisenstein. Projects include tying convex optimization to word embeddings to learn under hierarchical constraints as well as domain-knowledge grounding for end-to-end learning of effective representations of clinical text via deep learning.

Research and Development Intern, Sutter Health**May 2018 - December 2018**

Worked with the Research, Development, and Dissemination group and Professor Jimeng Sun (Georgia Tech) to develop deep learning methodology for disease prediction from clinical text.

Anderson Lab at the College of Charleston**January 2016 - May 2017**

Researched extensions to Google's Word2Vec algorithm used to generate word embeddings for variable-length documents. Investigated performance of the algorithm when used directly as a classifier, and whether this technique, along with similarly created ensemble methods, could outperform benchmark preprocessing and machine learning pipelines on topic recognition tasks. Presented as Bachelor's Thesis.

PROFICIENCIES

Daily Use	Languages: Python (Pytorch/Tensorflow /sklearn/pandas/nltk/numpy/multiproc), Bash. Tools: Git, TeX.
Past Use	Languages: R, Java, SQL, SAS, Octave. Tools: Dynet, Oracle RDBMS, MongoDB, Tableau.

SELECTED INVITED TALKS

“Quantifying Explanation in Neural Models.” NLP Seminar Series, University of Southern California (January 2020).

“Self-Attention for Universal Representations of Clinical Events”. Internship Presentation, Google AI (August 2019).

“**Machine Learning Approaches for Clinical Decision Making using Text**”. Data Science Stakeholders Meeting, Sutter Health (June 2018).

SERVICE & REVIEWING

NAACL 2021

BlackBoxNLP Workshop at EMNLP 2020

EMNLP 2019, 2020

ACL 2018 (subreviewer), 2019, 2020

AMIA Informatics Summit 2018, 2019

NeurIPS Machine Learning for Healthcare Workshop 2017, 2018, 2019

Publicity Co-Chair NAACL 2021

Student Volunteer EMNLP 2019, ACM FAT* 2019, NAACL 2018

AWARDS

NeurIPS Women in Machine Learning Workshop Travel Grant (2020). Funded virtual attendance; invited to present poster “Improving Neural Storytelling with Commonsense Inferences”.

Graduate Student Government Association Travel Grants (2019, 2020). Georgia Tech. Funded EMNLP, ACL, and NeurIPS attendances.

NeurIPS Travel Grant (2019). Funded attendance.

NeurIPS Women in Machine Learning Workshop Travel Grant (2019). Funded attendance; invited to present poster “Attention is not not Explanation”.

School of Interactive Computing Travel Grant (2019). Georgia Tech. Funded EMNLP attendance.

Computing Research Association (CRA-W) Grad Cohort for Women Attendee (2018, declined 2019/2020). Funded attendance at conference.

College of Computing Travel Grant (2018). Georgia Tech. Funded NAACL attendance.

Phi Kappa Phi Graduate Fellowship (2017). \$5,000 unrestricted.

Data Science Major of the Year, Departmental Honors (2017). College of Charleston.

Grace Hopper Scholar, the Anita Borg Institute (2016). Grace Hopper Celebration of Women in Computing Attendee (2015, 2016).

William Aiken Fellow (2013-2017). A fellowship representing the top 1% of students at the College of Charleston. \$92,000 towards tuition and fees.

Crosby Computer Science Award (2014). Awarded by professorial nomination to the most promising student in an introductory computer science course at the College of Charleston.

TEACHING AND SERVICE

Mentoring

Xiangyu (Becky) Peng, ML PhD Student, Georgia Tech (Fall 2020 - present).
Siyan (Sylvia) Li, undergraduate CS student, Georgia Tech (Fall 2020 - present).

Graduate Teaching Assistant, Georgia Institute of Technology

Deep Learning (graduate). **Spring 2021**
Deep Learning (graduate). ~200 students. **Fall 2019**
Guest Lecture: Transformers and Natural Language Applications.
Machine Learning (advanced undergraduate). ~100 students. **Spring 2019**

Women in Computing Club at the College of Charleston

2014 - 2017

President (2017), Vice-president (2016) and Treasurer (2015-16). Worked to promote diversity in computer science, host professional development workshops, and conduct community outreach.

EXTRACURRICULARS

Rock climbing (2017 - present)
Guitar (2004 - present)
Travel (have visited 25+ countries, 5 continents)
French (fluent at B2 level)