

# YOUNGSEO SON

www3.cs.stonybrook.edu/~yson · (631)997-9479 · ysson375@gmail.com

---

## Research Interests

---

My key research focus is in the field of Natural Language Processing (NLP) for social media analysis, language modeling, information extraction and data analysis. I collaborated with psychologists and computational linguists for human-centered language modeling to obtain higher accuracies of various NLP tasks from traditional tasks (e.g., sentiment analysis) to novel tasks such as discourse style analysis for psychological assessment and well-being measurement. I especially focus on discourse relation parsing to extract key information for targeted tasks such as opinion mining, detecting reasons for sentiment of reviews and political stance, or finding the correlations of discourse styles with human variables such as personality.

## Education

---

<b>STONY BROOK UNIVERSITY</b> , STONY BROOK, NY Doctor of Philosophy in Computer Science, GPA – 3.89/4.00	AUG 2015 – DEC 2020
<b>STONY BROOK UNIVERSITY</b> , STONY BROOK, NY <b>Summa Cum Laude, Honors Program</b> B.S Computer Science, GPA – 3.93/4.00	AUG 2013 – MAY 2015
<b>AJOU UNIVERSITY</b> , SUWON, SOUTH KOREA B.E Computer Engineering, GPA – 4.36/4.50 (Major GPA – 4.43/4.50)	MAR 2009 – AUG 2013

## Skills

---

- ML / AI / NLP / Deep Learning (7+ years of academia and industry experience): PyTorch, Tensorflow, Keras.
- Database: MySQL, Presto, Spark, MongoDB.
- Programming: Python, Java, MATLAB, C, C++, MFC, Javascript, JSP, XML, HTML, CSS, Hack.
- OS experience: Linux systems (managed Ubuntu GPU server), Mac OSX, and Windows.
- Documentation (Latex, MS Offices).

## Work Experiences

---

<b>Meta ML Research Scientist</b>	FEB 2021 - PRESENT
<ul style="list-style-type: none"><li>● Designed comment quality metrics and developed NLP models to recommend relevant comments and encourage high quality conversations (e.g., relevancy, sentiment and tones) (<a href="https://www.facebook.com/creators/engaging-audiences-building-community">https://www.facebook.com/creators/engaging-audiences-building-community</a>).</li><li>● Fine-tuned and deployed AI integrity models to mitigate hate speech on the platform (e.g., sexual, harassment and bullying) (<a href="https://ai.facebook.com/blog/the-shift-to-generalized-ai-to-better-identify-violating-content/">https://ai.facebook.com/blog/the-shift-to-generalized-ai-to-better-identify-violating-content/</a>).</li><li>● Conducted data science experiments and statistical analysis to develop ML models using state-of-the-art model architectures and optimize them for comment ranking across different products (e.g., Facebook public posts, Reels, live videos) (<a href="https://about.fb.com/news/2019/06/making-public-comments-more-meaningful/">https://about.fb.com/news/2019/06/making-public-comments-more-meaningful/</a>).</li></ul>	
<b>Facebook ML PhD Software Engineer Intern</b>	SUMMER 2020
<ul style="list-style-type: none"><li>● Worked as ML PhD Software Engineer (research) at Search Integrity team</li><li>● Developed graph-based ML and NLP model to improve search results</li><li>● Conducted data analysis and data mining to improve search models</li></ul>	
<b>Pacific Northwest National Laboratory (PNNL) Data Scientist PhD Intern</b>	SUMMER 2019
<ul style="list-style-type: none"><li>● Worked with Svitlana Volkova as my PI and Maria Glenski as my mentor.</li><li>● DARPA SocialSim Project: Developed SocialSim modules to analyze information / graph evolution and cross-platform misinformation / disinformation spreads on social media (e.g., Twitter, Reddit, Github)</li></ul>	

- Collaborated with Prasha Shrestha for detecting coordinated efforts and analyzing trends and spread mechanisms of cryptocurrencies over social media.

#### **Stony Brook University Research Assistant**

SUMMER 2017 - PRESENT

- Collaborating with Stony Brook Psychology and Stony Brook Medicine.
- Latent Dirichlet Allocation (LDA) Topics for content analysis of WTC responders' interviews and predicting health trajectories with languages.
- Developed a docx parser for the analysis using Differential Language Analysis Toolkit (DLATK).
- Correlating languages with human variables (e.g., personality, physical/mental health, income) by using language models including discourse relation parsing and topic clustering.

#### **World Well-Being Project (WWBP), the University of Pennsylvania, Research Scientist Intern**

SUMMER 2016

- Developed a discourse relation parser for social media to capture counterfactual thinking from tweets.
- Led the project with the help of Prof. Lyle Ungar and Anneke Buffone of the WWBP team.

#### **Dassault Systemes Korea Intern Software Engineer Internship**

DEC 2012 - FEB 2013

- Development of Product Lifecycle Management (PLM) Web Application (**ENOVIA**)
- Worked on updating PLM chart display and data visualization functions with PLM Development senior software engineers and sales team members

#### **Korean Augmentation to the US Army (Korean Army Sergeant)**

AUG 2010 - MAY 2012

- Worked as an Information Technology Specialist (25B).
- Worked with 2<sup>nd</sup> Infantry Division 8<sup>th</sup> Army of the United States deployed to South Korea.
- Managed and did troubleshooting on networks, computers, peripheral devices, and online portal of the battalion.

### **Publication**

Discourse Relation Embeddings: Representing the Relations between Discourse Segments in Social Media. **Youngseo Son**, Vasudha Varadarajan, H. Andrew Schwartz. In *The First Unimodal and Multimodal Induction of Linguistic Structures Workshop (UM-IoS) at the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.

Predicting Adolescent Depression and Anxiety from Multi-wave Longitudinal Data using Machine Learning. Mariah T. Hawes, H. Andrew Schwartz, **Youngseo Son**, Daniel N. Klein. In *Psychological Medicine (accepted in OCT 2022)*.

World Trade Center responders in their own words: Predicting PTSD symptom trajectories with AI-based language analyses of interviews. **Youngseo Son**, Sean A. P. Clouston, Roman Kotov, Johannes C. Eichstaedt, Evelyn J. Bromet, Benjamin J. Luft and H. Andrew Schwartz. In *Psychological Medicine 22 June 2021*.

Artificial Intelligence Language Predictors of Two-Year Trauma-Related Outcomes. Joshua R Oltmanns, H Andrew Schwartz, Camilo Ruggero, **Youngseo Son**, Jiaju Miao, Monika Waszczuk, Sean Clouston, Evelyn Bromet, Benjamin Luft, Roman Kotov. In *Journal of psychiatric research 143 (2021): 239-245*.

Author's Sentiment Prediction. Mohaddeseh Bastan, Mahnaz Koupaee, **Youngseo Son**, Richard Sicoli, Niranjana Balasubramanian. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*.

Suicide Risk Assessment with Multi-level Dual-Context Language and BERT. Matthew Matero, Akash Idnani, **Youngseo Son**, Salvatore Giorgi, Huy Vu, Mohammadzaman Zamani, Parth Limbachiya, Sharath Chandra Guntuku, H. Andrew Schwartz. In *NAACL 2019, CLPsych Workshop*

Causal Explanation Analysis on Social Media. **Youngseo Son**, Nipun Bayas, H Andrew Schwartz. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*

The Language of Well-Being: Tracking Fluctuations in Emotion Experience through Everyday Speech. Jessie Sun, H. Andrew Schwartz, **Youngseo Son**, Margaret Kern, Simine Vazire. In *Journal of Personality and Social Psychology*

Recognizing Counterfactual Thinking in Social Media Texts. **Youngseo Son**, Anneke Buffone, Anthony Janocko, Allegra Larche, Joseph Raso, Kevin Zembroski, H Andrew Schwartz, Lyle Ungar. In *ACL 2017*

Human Centered NLP with User-Factor Adaptation. Veronica E. Lynn, **Youngseo Son**, Vivek Kulkarni, Niranjan Balasubramanian, H. Andrew Schwartz. In *EMNLP 2017*

Capturing the Human Nuances of Optimism using Discourse Relations and Explanatory Style Recognition. **Youngseo Son**, Nipun Bayas, H Andrew Schwartz. In *MASC 2017*

Health Prediction based on Optimistic Explanatory Styles in Social Media. **Youngseo Son**, Laura Smith, H Andrew Schwartz. In *MASC 2016*

Predicting Users' Features from Their Relationships with Other Users. **Youngseo Son**, Yanhong Liu. In *Stony Brook URECA 2015*

## Research Projects

---

**9/11 World Trade Center Oral History Interviews for PTSD Symptom Analysis** AUG 2017 – OCT 2021  
(Project in collaboration with Stony Brook WTC Wellness Program and Stony Brook Medicine)

- Analyzing interviews of people who were at the scene of 9/11 WTC Attack
- Correlating linguistic features of the subjects with their mental/physical health
- Using LDA topic clustering, discourse relation parsing, sentiment/emotion lexicons

**Suicide Risk Assessment on Reddit Posts** MAR 2019  
(HLAB collaboration for NAACL SharedTask)

- Ranked No 1. for predicting reddit users' suicide risk level using their SuicideWatch and Non-SuicideWatch posts (Task B)
- Developed user-factor-adapted RNN models with post-level attention using BERT and psychology language model representations of reddit posts

**The Language of Well-Being Project** JUL 2017 – DEC 2018  
(Project in collaboration with the University of California, Davis and the University of Melbourne)

- Correlating linguistic features of people's everyday language with the changes of their emotions
- Conducting LDA topic clustering over the transcripts of the participants' daily speech for the emotion analysis
- Using N-gram, Linguistic Inquiry and Word Count (LIWC), sentiment/emotion lexicons

**Health Prediction based on the Explanatory Style Analysis of Social Media** AUG 2015 – NOV 2018  
(Project in collaboration with the University of Pennsylvania and Prof. Schwartz)

- Presented at Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2016 and 2017)
- Using state-of-the-art discourse dependency parsers such as RST parsers or PDTB parsers for distracting discourse relations from messages collected from Facebook
- Analyzing Facebook status updates in order to predict users' personality and mental health.
- Data collection using Amazon Mechanical Turk
- Bidirectional LSTM for PDTB-style discourse relation parsing (causality) in social media texts

- Personality Analysis Using Linguistic Styles of Social Media Languages** MAY 2016 – AUG 2017  
 (Project in collaboration with the University of Pennsylvania)
- Leading the project with help of Prof. Ungar and Anneke Buffone of the World Well-Being Project
  - Analyzing correlations between syntactic structures of Facebook Status updates and profile information, personality, socio-economic statistics of users
  - Worked as a Summer Intern Researcher in the Positive Psychology Center of the University of Pennsylvania

- Privacy Policy Language Modeling using Prolog** AUG 2014 - May 2015  
 (Senior Honors Research Project under Advisor Professor Y. Annie Liu & Logic Programming Group of Stony Brook)
- Facebook privacy policy translation for Object-Oriented Modeling and XSB, Flora for the Relational Modeling
  - Creation of automatic policy conflict detector aiming to prevent discrepancies between Facebook privacy policy and the terms of use of Third-Party partners

- User Profile Prediction Based on Relationships in Social Networks** SUMMER 2014  
 (Undergraduate Research & Creative Activities (URECA) 2014, Stony Brook University)
- Users profile information prediction from their relationships with other users in Google+, Facebook, and Twitter dataset from Stanford Network Analysis Project
  - Outlier Detection to social network relationships with Local Outline Score from “Oddball”
  - Published in Stony Brook URECA Collected Abstracts and made a presentation for the Celebration of URECA Symposium

## Teaching Experience / Activities

- Conference on Empirical Methods in Natural Language Processing (EMNLP) Reviewer**  
**Annual Meeting of the Association for Computational Linguistics (ACL) Reviewer**  
**Stony Brook University Computer Science TA** SPRING 2014 – FALL 2019
- Teaching Assistance for Big Data Analytics, Senior Software Engineering Project, Computer Science III (Java Programming), Advanced Game Programming, and Computer Music
- Elementary School Science Lab Teacher (Graduate Student Volunteers)** SPRING 2016 – PRESENT
- Give special topic science labs to the fifth graders of the Nathaniel Woodhull Elementary School
- Stony Brook University Korean Graduate Students Association Vice President** FALL 2017  
**Korean-American Scientist and Engineers Association (KSEA) Volunteer** FALL 2016, FALL 2017  
**Ajou-Stony Brook University Students Association President** SPRING 2015  
**Stony Brook Computing Society** FALL 2013 – PRESENT

## Awards / Scholarship

- Special CS Department Chair Fellowship AUG 2015  
 Stony Brook University Computer Science Award of Honor MAY 2015  
 Stony Brook University URECA Stipend SUMMER 2014  
 Stony Brook University Outstanding Academic Achievement Awards FALL 2013 - SPRING 2014  
 Stony Brook University Dean’s List Nomination FALL 2013 - SPRING 2014  
 Ajou University Superior Academic Performance Scholarship FALL 2009 - SPRING 2013  
 US Army Best Korean Augmentation to the United States Army (KATUSA) of 2012 MAY 2012  
 US Army Best Warrior 2012 APR 2012