

# Jon Saad-Falcon

jonsaadfalcon@gmail.com | (912) 713-0720 | [LinkedIn](#) | [Github](#) | U.S. Citizen

## Education

**Georgia Institute of Technology (Georgia Tech)** - Atlanta, GA

August 2018 - May 2022

*B.S. - M.S. in Computer Science*

**GPA: 3.96**

- **Minors:** Mathematics and Linguistics
- **Study Abroad:** Berlin Summer Program at the Technische Universität Berlin (Summer 2019)
- **Undergraduate Thesis:** PeopleMap: NLP and Visualization Tool for Mapping Out Researchers
- **Scholarships:** Stamps President's Scholarship (premier full-ride scholarship)

---

## Research Experience

**Allen Institute for Artificial Intelligence (AI2)**, Semantic Scholar, Seattle, WA

*August 2021 - Present*

*Predoctoral Young Investigator with Daniel Weld, Tom Hope, and Jonathan Bragg*

- Develop semantically-logical tokenizers for information extraction of biomedical documents
- Create relation extraction techniques for developing knowledge graph of biomedical entities
- Evaluate new prediction techniques for recommending papers before citation and readership data become available

**Stanford University**, Center for the Study of Language and Information (CSLI), Palo Alto, CA

*June 2021 - August 2021*

*Research internship with Christopher Potts*

- NSF-Funded REU opportunity at the Stanford Artificial Intelligence Laboratory (SAIL)
- Constructed new benchmark dataset for naturalistic, technical question-answering models
- Developed open-resource generation script for building technical datasets using forum websites

**Allen Institute for Artificial Intelligence (AI2)**, Semantic Scholar, Seattle, WA

*October 2020 - May 2021*

*Research internship with Daniel Weld and Tom Hope*

- Constructed search engine for exploring challenges and future directions within COVID-19 literature
- Collaborated with 10+ medical professionals to design better tools for assisting scientific research
- Built clustering algorithms for accelerating the rate of hypothesis generation and literature-based discovery

**Natural Language Processing Research**, Georgia Tech, Atlanta, GA

*August 2019 - Present*

*Research assistant with Diyi Yang*

- Developed project to computationally analyze persuasion strategies and contexts within dialogue datasets
- Extracted rhetorical strategies from persuasive requests to find patterns and connections in successful attempts
- Analyzed LinkedIn datasets using LSTM models to understand and predict career paths

**Human-Centered Artificial Intelligence Research**, Georgia Tech, Atlanta, GA

*August 2019 - Present*

*Research assistant with Polo Chau*

- Led project to build visual, explorable map of researchers and their work using textual embedding techniques
- Explored strategies for predicting and visualizing energy consumption in deep learning models using PyTorch and Svelte.js
- Constructed open-source, graph-based application for scientific literature exploration used by over 1000 students

---

## Publications

[1]: *A Search Engine for Discovery of Scientific Challenges and Directions*

Dan Lahav, **Jon Saad-Falcon**, Bailey Kuehl, Sophie Johnson, Sravanthi Parasa, Noam Shomron, Polo Chau, Diyi Yang, Eric Horvitz, Daniel Weld, and Tom Hope

*AAAI 2022*

[2]: *ColBERTv2: Effective and Efficient Retrieval via Lightweight Late Interaction*

Keshav Santhanam\*, Omar Khattab\*, **Jon Saad-Falcon**, Christopher Potts, Matei Zaharia

*Preprint, under review, arXiv:2112.01488, 2021*

[3]: *Argo Scholar: Interactive Visual Exploration of Literature in Browsers*

Kevin Li, Haoyang Yang, Anish Upadhayay, Zhiyan Zhou, **Jon Saad-Falcon**, Polo Chau

*IEEE VIS 2021*

**VAST Best Poster Research Award, Honorable Mention**

[4]: *Quantifying the Impact of Human Capital, Job History, and Language Factors on Job Seniority*  
Austin Wright, Caleb Ziem, Haekyu Park, **Jon Saad-Falcon**, Polo Chau, Diyi Yang, Maria Tomprou  
*Preprint, under review, arXiv:2106.11846, 2021*

[5]: *EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models*  
Omar Shaikh, **Jon Saad-Falcon**, Austin Wright, Nilaksh Das, Scott Freitas, Omar Asensio, Polo Chau  
*ACM CHI 2021*

[6]: *Examining the Ordering of Rhetorical Strategies in Persuasive Requests*  
Omar Shaikh, Jiaao Chen, **Jon Saad-Falcon**, Polo Chau, Diyi Yang  
*Findings of EMNLP 2020*

[7]: *Mapping Researchers with PeopleMap*  
**Jon Saad-Falcon**, Omar Shaikh, Zijie Wang, Austin Wright, Sasha Richardson, Polo Chau  
*IEEE VIS 2020*  
**VAST Best Poster Research Award, Honorable Mention**

## Honors and Awards

<b>Donald V. Jackson Fellowship</b> , Georgia Tech	Spring 2021
<b>President's Undergraduate Research Award</b> , Georgia Tech	Fall 2020
<b>Computer Science Research Mentorship Program</b> , Google	Spring 2021
<b>D.E. Shaw Nexus Fellowship</b> , D.E. Shaw and Co.	Fall 2020
<b>U.N. Millennium Fellowship</b> , United Nations	Fall 2021
<b>Tapia Scholarship</b> , ACM Richard Tapia Conference	Spring 2020
<b>1st place, Citadel Securities Trading Challenge</b> , Citadel Securities and Georgia Tech	Spring 2020

## Industry Experience

**Goldman Sachs**, Global Investment Research, New York City, NY *July - August 2020*  
*Summer Research Analyst*

- Initiated independent project using quantitative techniques and NLP tools to perform valuation of companies
- Collaborated with traders and external clients to develop new automated trading strategies
- Assisted energy sector research team as they advise corporate clients on financial decisions

**Etude LLC**, Atlanta, GA *August 2018 - March 2020*  
*Co-founder of student technology startup*

- Led development of Étude, an advanced PDF reader designed for professionals
- Established partnership with Wolfram Research to implement their natural language processing APIs in a new product
- Awarded 1st Place at Georgia Tech's Inaugural Catapult Start-up Competition

## Conference Participation

**EMNLP 2020**, Student Volunteer - November 2020  
**ACL 2020**, Student Volunteer - July 2020

## Skills

**Programming:** Python, Java, and Javascript (advanced), Scala (intermediate), C and Assembly (proficient)  
**Software:** PyTorch, TensorFlow, Keras, Apache Spark, Svelte.js, React.js, Node.js  
**Communication:** presentations, conference papers, technical reports, lab write-ups, research posters  
**Spoken Languages:** English (native), Spanish (native), Arabic (beginner), French (beginner)