ANA SMITH

https://analysmith.com · alsmith@cs.cornell.edu

Seeking **full time employment or post-doc** working as an **NLP researcher**

EDUCATION AND FELLOWSHIPS

Cornell University → **Department of Computer Science** → Professor Lillian Lee (advisor)   *August 2016 - Present*
PhD Student (Expected graduation: May 2023)
M.S. Computer Science (conferred May 2020)

Fulbright Research Fellowship (South Korea)   *August 2015 - June 2016*
Georgia Institute of Technology   *August 2011 - May 2015*
B.S. Computer Science, minor in Korean Language

SELECT DISSERTATION PROJECTS

Information Extraction for Modeling SCOTUS Oral Arguments To be submitted to ACL Rolling Review.

War in Piece: Comparing Perspectives in Wikipedia Presented at Text-As-Data 2022. Accepted at SIGHUM'22.

Investigating Policy Invocations in Featured Article Candidate Nominations Under revision for ICWSM'22.

RECENT INDUSTRY EXPERIENCE

The New York Times   Ithaca, NY (June 2020 - Aug 2020)
Data Science Intern (remote), Manager: Zhen Yang, Team Lead: Anna Coenen
- Fine-tuned large language models for algorithmic recommendations team
- Modified pre-processing pipeline for Latent Dirichlet model using Gensim
- Prepared interactive demos for data science team to explain fine-tuning methods using Pytorch / HuggingFace

Duolingo   Ithaca, NY (June 2020 - Aug 2020)
Research Intern (remote), Mentor: Klinton Bicknell
- Implemented machine learning models for aiding learner progress
- Proposed and implemented semi-supervised method for forum data organization and processing
- Conducted A/B experiments and analysis for improved product experience

IBM Research   Yorktown, NY (May 2019 - Aug 2019)
Research Intern, Mentors: Song Feng and Q. Vera Liao, Manager: Luis Lastras
- Trained dialogue models pretrained transformer language models (BERT and DistilBERT)
- Created conversation annotation tool with Flask web framework for internal data labeling
- Conducted dialogue modeling results, analyzed results, and wrote paper submitted to ACL.
- Jointly filed an application for patent (United States)

Pindrop Security   Atlanta, GA (May - Aug 2014, May - Aug 2015, Jun - Aug 2017)
Research & Development Intern (Some results cannot be discussed due to company policy)
- Implemented scrapers, applied machine learning methods, and conducted analysis for research engineering team
- Scraped and validated caller complaints in order to identify scam patterns
- Designed features and experiments for the improvement of fraud detection

SKILLS & SAMPLE CODE

Github: https://github.com/analysmith
Platforms / Libraries / IDEs (alphabetical): Anaconda, AWS, BeautifulSoup, Google Cloud, HuggingFace, JSON, Jupyter, kenLM, LaTeX, mwparserfromhell, numpy, pandas, ParlAI, PyCharm, PyTorch, scipy, scikit-learn, seaborn, spaCy, Stanford CoreNLP, virtualenv
Programming experience (in order of preference): Python (primary), MySQL, Java, Flask, JQuery, JSON, HTML/XML, CSS, Jinja2, Javascript, GWT, ActionScript, MATLAB, C
Languages: English (native speaker), Korean (conversational)