Yao Dou

CODA Tech Square 1183R, Atlanta, GA • (206)-379-5769 • douy@gatech.edu

E D U C A T I O N

Georgia Institute of Technology, Atlanta, GA
Ph.D. in Computer Science   2021 – present
Advisor: Prof. Wei Xu

University of Washington, Seattle, WA
B.S. in Computer Science   2017 - 2021

- **GPA:** 3.89/4.0
- **Extracurricular Activities:** Co-founder of the UW chapter of Triangle Fraternity
- **Honors and Awards: Microsoft Endowed Scholarship** 2020 – 2021; **Dean's List** 2017 - 2021; **Best Analysis Project** in INFO 201 Class

E X P E R I E N C E

Research Intern
Allen Institute for AI, January 2021 – March 2021
With Prof. Yejin Choi

- Did Scarecrow project and wrote the paper

Research Assistant
University of Washington xlab, April 2019 – July 2021
With Prof. Yejin Choi

- Did Scarecrow project and wrote the paper
- Did MultiTalk project and wrote the paper
- Presented paper in lab meeting

Data Science Intern
Noonum, INC, June 2020 – September 2020

- Finetuned Spacy-large model, BERT and Roberta model to do Named Entity Recognition
- Finetuned BERT embeddings to better capture semantically similarity in finance domain
- Introduced a feature about related concept suggestion and deploy it on Noonum dashboard

Teaching Assistant
CSE446 (Machine Learning) at University of Washington, Spring 2020, Fall 2020
With Prof. Kevin Jamieson and Prof. Jamie Morgenstern

- Designed assignment question and quiz section worksheet
- Held weekly office hour and answered online questions
- Attended meetings and graded assignments

Research Assistant
University of Washington Autonomous Flight Systems Laboratory, January 2018 – September 2018

- Flight tested aircrafts and ensured they are tuned appropriately for safe and reliable flight
- Formatted maintenance files of all drones
- Improved interface for the LIDAR sensor

P R O J E C T S

Scarecrow: A Framework for Scrutinizing Machine Text
Yao Dou, Maxwell Forbes*, Rik Koncel-Kedziorski, Noah A. Smith, Yejin Choi*
https://arxiv.org/abs/2107.01294
July 2020 – July 2021

- Developed a new structured, crowd-sourced error annotation scheme called Scarecrow
- Collected over 41k spans on human and machine generated text
- Quantitatively and qualitatively analyzed the text generation systems in terms of Scarecrow on news text generation
- Submitted to **ACL Rolling Review**

MultiTalk: A Highly-Branching Dialog Testbed for Diverse Conversations

Yao Dou, Maxwell Forbes, Ari Holtzman, Yejin Choi

https://arxiv.org/abs/2102.01263

April 2019 – June 2020

- Collected a large dataset of highly-branching written conversations
- Introduced a scoring algorithm based on the weighted bipartite graph matching
- Computed metrics like BLEU, did sentiment analysis, emotional and dialogue act tagging on the dataset and generations
- Finetuned the GPT2 model to generate sentence that can lead to a future response with an intended emotion
- Accepted to **AAAI 2021**

Exploration of Seattle's housing prices

https://uw-cse442-wi20.github.io/FP-seattle-housing-price/

February 2020 – March 2020

- Built an interactive tool that find desired district of housing by user input preference list
- Analysis the data of Seattle's housing prices, traffic, crime stats, schools, stores, restaurants, companies and transportation
- Did data visualization about the relationship between housing prices and other factors such as traffic, crime stats and etc.
- Collaborated with 3 team members

Interactive Shiny Based Stock Analysis Application

February 2018 – March 2018

- Project awarded as the **Best Analysis Project**
- Led a 4-person team to implement an interactive Shiny based application which including functions: visualizing stock data of S&P 500, comparing different companies' stocks and predicting stock in a period
- Collaborated with 3 team members

Building Alexa Skills Kit to voice control a thermostat

December 2017 – January 2018

- Developed a custom Alexa skill to control the thermostat with multi-turn interactions
- Used ESP8266 to record the temperature and insert it into a Cloud database periodically
- Alexa reads or changes data from the database according to the instruction given by people to Alexa

NOTABLE SKILLS

- **Computer:** Python, Java, C++, R, HTML, JavaScript, jQuery, CSS, PyTorch, AWS
- **Languages:** Chinese (Native), English (Fluent), French (Elementary)