Sophia Folino

(475) 206 – 1528 | srf77@cornell.edu | LinkedIn

Experience

Natural Language Processing (NLP) Intern – TransPerfect, New York, NY | June 2023 – August 2023
Internship with an Artificial Intelligence (AI) team building machine translation solutions in Python.

- Conducted exploratory data analysis on translation memory to uncover errors in pre-processing phase.
- Identified and cleaned contamination across training and test sets to improve model evaluation.
- Wrote scripts to measure text similarity for translation evaluation using Levenshtein distance, BLEU scores, chrF scores, COMET, TF-IDF scores, N-gram (perplexity) scores, and Cosine similarity scores.
- Designed a test set to better represent a client domain using clustering algorithms to organize data. Used bag of words, TF-IDF and sBERT vectorization and an N-gram language model.
- Used client data to perform domain adaptation on a Canadian French to English transformer model and the Marian toolkit to finetune the transformer model for hyperparameter optimization.

Teaching Assistant

- *Advanced NLP for Humanities Research*, mentoring students in a project course (Spring 2024).
- *Foundations of Artificial Intelligence*, covering AI topics including NLP and machine learning (Fall 2023).
- *Language and Information*, covering analysis of unstructured textual information (Spring 2023).
- *Fundamental Programming Concepts*, covering basic programming and data structures (Summer 2022).

Facilitator – Academic Excellence Workshops
Taught weekly workshops for *Object-Oriented Programming & Data Structures* (Fall 2022), *Computing Using Python* (Spring 2022), and *Calculus 2* (Fall 2021).

Education

Cornell University – College of Engineering, Ithaca, NY | Fall 2020 – Present
Bachelor of Science in Computer Science expected May 2024. GPA: 3.5.

- **AI – Language Specialization:** Machine Learning, Reinforcement Learning, Large Language Models, AI Practicum, Natural Language Processing, Computational Linguistics, Intro. to Linguistics, Italian 1 & 2.
- **Computer Science Core**: Introduction to Computing, Object-Oriented Programming & Data Structures, Data Structures & Functional Programming, Computer System Organization, Operating Systems, Algorithms.
- **Mathematics**: Calculus 2 & 3, Discrete Math, Linear Algebra, Probability & Statistics, Operations Research.
- **Programming Languages & Tools**: Python, Java, OCaml, C, PyTorch, NumPy, Pandas, AWS Cloud.

Projects

- **AI Practicum:** Developed a "Codenames" game using machine learning to generate clues based on semantic similarity and to learn from incorrect guesses.
- **NLP:** Implemented multiple models for named entity tagging: Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM), Feed-Forward Neural Network (FFNN), and Recurrent Neural Network (RNN). Used Long Short-Term Memory (LSTM) and Encoder-Decoder models for semantic role labeling and trained a large language transformer model to generate humorous captions for cartoons.
- **Functional Programming:** Developed Enigma encryption software, a text adventure game engine, a red-black tree search engine, a Robot Meta-Language Interpreter, and a Wordle-based game in OCaml.

Other

- Lived in Denmark and studied in a Danish High School, 2016 – 2017. Hold U.S. and Danish citizenship.
- Runner, reader, and dedicated Shiba Inu dog owner.