Joseph Cappadona

(516) 220-2837 | josephcappadona27@gmail.com | GitHub: josephcappadona | https://jcappadona.com

Education

New York University, M.S. Computer Science 2021 – 2023
GPA: 3.97

University of Pennsylvania, B.S.E. Computer Science 2014 – 2018

Professional Experience

Deep Learning Engineer, NYU Langone September 2022 – Present
- Implemented self-supervised learning (SSL) pipeline for 3D breast cancer MRIs, including distributed training over 40+ GPUs, experiment tracking, and visualization
- Fine-tuned various SSL pre-trained models on a breast cancer diagnosis task and demonstrated that fine-tuning the SSL pre-trained models yields a 1% absolute improvement relative to fine-tuning a Kinetics-400 pre-trained model

Machine Learning Engineer, XaiPient July 2019 – June 2020
- Implemented and maintained library for Explainable Artificial Intelligence (XAI), wrote corresponding documentation
- Implemented algorithms encompassing attribution methods, adversarial examples, rule-based explanations, counterfactual explanations, and fairness/bias analysis
- Applied XAI library to finance and healthcare datasets, designed visualizations, and presented them to potential clients

Machine Learning Projects

Language Models are Unreliable Spatial Reasoners github.com/josephcappadona/spatialQA
- Used procedural generation to create natural language inference (NLI) sentence pairs that test four categories of spatial reasoning: motion, orientation, distance, and containment
- Tested performance of seven state-of-the-art NLI and question-answer (QA) models (e.g., GPT-3, RoBERTA, DeBERTa) on the generated test suite
- Found that no model achieved better than 80% on any category, with all models averaging below 65% across categories, demonstrating that all state-of-the-art language models lack very fundamental spatial reasoning abilities

Aggregating Insights from Amazon Product Reviews github.com/ryin1/senior-design-f17
- Used word2vec to create embeddings for product features based on Amazon product reviews
- Clustered embeddings in order to group synonymous features and bootstrapped domain-specific opinion words based on part-of-speech and syntactic dependency information
- Computed sentiment scores for product feature clusters and displayed the results in a web application

Data Engineering Projects

Kids Britannica Dataset github.com/josephcappadona/kids_britannica
- Used BeautifulSoup to scrape kids.britannica.com for three tiers of multimodal articles (125M tokens over 130k articles) and analyzed the relationship between sentence-level statistics and article tier

Jeopardy QA Dataset github.com/josephcappadona/j_archive
- Used BeautifulSoup to scrape j-archive.com in order to construct a QA dataset

Technical Skills

Languages: Python, Java, C/C++, JavaScript, Swift/Objective C, Bash, HTML/CSS, LaTeX

Libraries: PyTorch, Tensorflow, Hugging Face, NumPy, Pandas, scikit-learn, Spacy, OpenCV, Matplotlib, Seaborn, D3, BeautifulSoup, Scrapy, Selenium, Requests, Flask, React

Extracurricular

Hobbies: Chess (NYU Chess Team President), Poker, Yoga, Rock Climbing, Hiking