Andrew Lyubovsky

a.l.andlyu@gmail.com linkedin.com/in/andrew-lyubovsky scholar.google.com/citations?user=ZmideGIAAAAJ

Education

Carnegie Mellon University - School of Computer Science Pittsburgh, PA
M.S. in Intelligent Information Systems (ML + NLP) (GPA: 4.03/4.00) Dec 2022
Courses: Multimodal ML, ML for Text Mining, Deep Reinforcement Learning, Cloud Computing

College of William & Mary Williamsburg, Virginia
Bachelors of Science in Computer Science and Neuroscience (Double Major) (GPA: 3.83/4.00) May 2021
Courses: Adv. Linear Algebra, Adv. Prob Stats, Computational Neuroscience, Philosophy of Technology

Experience

Amazon, AWS Santa Clara, CA
Applied Science Intern | Machine Learning Solutions Lab May 2022 - August 2022

- Developed a generalizable information extraction (IE) model by formulating IE tasks in an extractive **question-answering** format and evaluated zero and few-shot performances to ensure robustness across diverse customer needs.
- Created a robust 5-fold cross-validation framework to assess performance on unseen classes to establish a rigorous benchmark.
- Developed a pipeline in **SageMaker** utilizing a pre-trained **RoBERTa** model on the SQuAD Dataset, resulting in an improvement of up to 30% over the **BERT-base** model.

Directed Research

Comic Frame Image Generation MIIS Capstone Project | Sept 2022- Dec 2022
- Developed a novel zero-shot **image-to-image** generation approach to introduce concepts from selected images into generated images to ensure character consistency between multiple frames when generating a comic.
- Ported **Pytorch** infusion method to **Flax** to improve training speeds fifteen-fold when running the model on **TPUs**.

Neurosymbolic Reasoning in NLP under Dr Yiming Yang MIIS Directed Study | Sept. 2021- May 2022
- Surveyed papers and researchers to formulate Neural-Symbolic Reasoning as a method of embedding external knowledge in the form of structured data into neural networks, improving interpretability and robustness.
- Explored multiple "reasoning" methods such as code and logic generation that "reason" over inputs, **Graph CNN** that reason over graph edges, and the use of Commonsense datasets that embed "reasoning" into neural networks.
- Workshop Paper: "Lyubovsky, A., Madaan, A., Yang, Y., et. al. "Characterizing Neuro-Symbolic Reasoning in NLP." SUKI (2022).

William Mary Computer Science | Research Experience for Undergraduates (REU) June 2020 - May 2021
- Developed model to predict game injuries based on sensor data collected from footaball players over the course of a season.
- Benchmarked several classification algorithms (**SVM**, **Random Forest**, **Logistic Regression**) to evaluate the predictive power of various custom features and the importance of sensors for football injury prediction (through statistical significance test on **5-fold cross-validation**.)
- Journal Paper: "Lyubovsky, A., Liu, Z., Watson, A., et. al. "A Pain Free Nociceptor: Predicting Football Injuries with Machine Learning." Smart Health 21 (2021).

Project Highlights

Video Question Answering Jan 2022 - May 2022
- Improved the performance of VideoQA model (**HCRN**) by minimizing the L2 distance between features of consecutive frames for the **Action Genome Question Answering** (AGQA) benchmark.
- Formulated a two-stage VideoQA pipeline comprising (1) a captioning module using **GPT-2** fine-tuned to generate captions solely using image features (**ResNet**) and (2) a **seq2seq model (T5)** that conditions the caption to generate the answer.

IdeaNet (ideanetapp.com) Dec 2021-present
- Developed a semantic search engine using a **BERT-based** sentence transformer to retrieve ideas similar to a user query and to connect people with similar ideas within an organization.
- Conducted user interviews to understand the challenges in identifying potential teammates, and forming teams after an initial contact occurs between potential teammates.
- Collected experimental data (student ideas) by scraping Google Scholar and summarizing abstracts with **GPT3**.
- Deployed **Django API** using **AWS Elasticbean** on **EC2** and an **RDS** database to run a pilot study, matching students with mentors for AI mentor-match program at CMU.

Gymnastics

William Mary Mens Gymnastics 2018 - (2021 Sad face)
- Practiced 20 hours per week, Organized training plan with coaches, Contributed to team reinstatement, Lead team by example.
- **Awards**: 2021 NCAA All-American, WM Record Holder on Highbar, CoSIDA Academic All-American.

Technical Skills

Languages: Python, Numpy, Pytorch, Pandas, Matplotlib, Scipy, Keras, OpenCV, Tensorflow, Java, Bash, C, HTML/CSS

NLP: Huggingface, LSTMs, Tokenization, Text Generation, BERT, Visual/Video Question Answering

Other: Git, AWS, GCP, Sagemaker, Docker, Kubernetes, ETL, Spark, Load Balancing, MySQL, Linux, Bash, Signal processing