Yang Ouyang

Durham, NC 27705, United States
(984) 325-2686 yang.ouyang@duke.edu LinkedIn GitHub

Education Experience

Duke University **Aug. 2022 – May 2024**
Master of Engineering in Electrical and Computer Engineering *Durham, U.S.A*

- GPA: 4.0 / 4.0
- **Teaching Assistant** of ECE 551K: Programming, Data Structures, and Algorithms in C++

Shenzhen University **Sep. 2018 – July 2022**
Bachelor of Engineering in Computer Science and Technology *Shenzhen, China*

- GPA: 3.75 / 4.0
- Honors/Awards: Two times winner of The Second Award of Studying Star in 2020 & 2021 (Ranked in 4 & 6);

Research Projects

Adversarial Privacy Attacks on Aligned Large Language Models **Oct. 2023 – Present**
Durham, U.S.A

- Conducted comprehensive experiments using Greedy Coordinate Gradient to identify exact privacy leakage (**90%** and even all of the **Output** from material) without directly related prompt of Large Language Models like StableLM-Tuned-Alpha and StableVicuna-13B which are fine-tuned using Reinforcement Learning from Human Feedback on various conversational and instructional datasets.
- Subsequent to this analysis, we explored two avenues: developing robust countermeasures to reinforce model privacy or innovating an enhanced adversarial approach to refine the RLHF training protocol, thereby mitigating potential privacy exploitation.

Enhancing Pre-trained Data Detection for LLM Privacy Protection **Jan. 2024 – Present**
Durham, U.S.A

- Refined the MIN-K% PROB metric using temperature scaling, achieving a 5% improvement over benchmark methods for detecting pre-trained data in LLMs.
- Developed a novel gap-based method (GAP) for pre-trained data detection, improving AUC by 10% over the state-of-the-art (MIN-K% PROB) by measuring log probability density gaps within datasets.

Addressing Data Scarcity in Multimodal Models Jan. 2024 - Present at Durham, U.S.A **Jan. 2024 – Present**
Durham, U.S.A

- Developing methods to generate high-quality synthetic multimodal datasets. Leveraging ChatGPT 4 for in-context prompt generation, driving image creation with Stable Diffusion models, and employing techniques like BoxDiff for precise image-text alignment.

Relaxing Crack Scarcity: Data Augmentation for Imbalanced Crack Recognition **July 2022 – Nov. 2022**
Shenzhen, China

- Synthesized diverse crack samples in the feature space by disentangling and reassembling crack-relevant and irrelevant features, effectively augmenting data to alleviate class imbalance.
- RELAX notably improved crack class recognition by approximately 9% in the INPP2022 dataset, with a minimal performance drop in the majority class.

Internship Experience

Trip.com Group Ltd | *Java, Spring Framework* **May 2023 – Aug. 2023**
Back End Developer Intern, Flight Ticket Department *Shanghai, China*

- Contributed to the optimization of MegaSearch which serves as an aggregation and cache layer for Trip's international ticket responses using **Java**.
- Optimized the response size to fit AWS's smaller bandwidth while saving some storage costs. Reduced the **Protobuf** response size by 50% in total using a variety of methods.
- Compared a variety of serialization and deserialization means using **JMH**: including the latest open source Fury, Kryo, and ultimately found that Protobuf is the most efficient serialization, but Kryo in the serialization of the size of a small advantage.

Amazon Web Service | *Java, K8s* **July 2022 – Oct. 2022**
Back End Developer Intern, DeepJavaLibrary Department *San Jose, U.S.A(remote)*

- Integrated the DeepJavaLibrary Model Server with the open-source KServe platform deeply through a well-thought-out plan.
- Developed 3 HTTP APIs applicable to the KServe inference engine for DJL-Serving using **Java**, which respond to the users with the DJL-Serving running model's health status, the serving model's information, and inference results which also need the request data.
- Made each API return a response code and pass the corresponding unit test.
- Hosted containerized DJL-Serving on KServe, writing **yaml** files specifying its ports, and related parameters.
- The specified DJL-Serving model can be run in the KServe framework by deploying a test **yaml** file.

Tencent Music Entertainment Group | *Javascript, Vue* **May 2021 – Sep. 2021**
Front End Developer Intern, Security Center *Shenzhen, China*

- Applied **Vue2.0** framework based on JavaScript to develop the inner front-end of content audit security platform.
- Built and maintained middle ground management system.
- Developed search, collection, and recently used functions for the middle ground management system.
- Utilised Least Recently Used (LRU) to design a cache that was able to clear the cache efficiently.
- Configured Webpack to optimize the local development and deployment increased the packaging speed by 75% and decreased the packaging size by 10%.

Technical Skills

- **Programming Languages**: Java, Python, C, C++, JavaScript
- **DeepLearning Frameworks**: PyTorch