Zinnia Nie

zinnia.nie@gmail.com · 774-423-9940
LinkedIn: https://www.linkedin.com/in/zinnia-nie-468850224/ · GitHub: https://github.com/zinnia11

Education

University of Wisconsin-Madison | Sept 2021 - Dec 2024
Bachelor of Science: Computer Science and Data Science | GPA 3.95/4.00

Certificate: Game Design
Languages: English (fluent), Mandarin Chinese (fluent)
Relevant Honors:

- Wisconsin Merit Scholarship
- William F. Vilas Merit Scholarship 2023-2024
- Dean's List Fall 2021 - Fall 2024

National University of Singapore | Jan - May 2024
Non-Graduating Programme in the School of Computing (Study Abroad/Exchange student)

Experience

WISCERS Spring 2024 Cohort | 1/2024 - present
Mentor: Ramya Korlakai Vinayak

- Researched methods in Out of Distribution (OOD) detection with guidance from a graduate student mentor
- Participated in a project about pluralistic alignment in large language models
  - Ran experiments using a subset of the Jester dataset looking for heterogeneous preferences
- Co-authored a paper on the long tail distribution of culturally significant items in image generation models
  - Built inference pipelines using HuggingFace and PyTorch for text-to-image models like SDXL, FLUX.1, SD1.5, and SD3.5 Large
  - Generated, graphed, and evaluated CLIPscores and VQA scores
  - Evaluated img-img similarity between ground truth images and generated images of the category
  - Identified geographic biases in the text-to-image models and created a new metric to measure the long-tailedness of cultural items and regions
  - Launched a user study to identify perceptual judgements and potential offensive content from the T2I systems

Student Analyst | 4/2022 - 12/2023
UW-Madison Cybersecurity Operations Center

- Performed daily tasks such as email header analysis, malicious domain research, and threat hunting to maintain device and data security and protect the populace of UW-Madison
- Learned about computer networks and cybersecurity weaknesses, attacks, and preventative methods
- Communicated with a team of other student analysts to complete daily tasks efficiently
- Provided feedback about processes to support continuous process improvement

Business Intelligence Intern | 5/2023 - 8/2023
Spectrum Brands, Inc.

- Developed a performance testing methodology on business reporting dashboards deployed in SAP Analytics Cloud and serviced by SAP HANA to be used in the future as part of the dashboard creation process
- Implemented said performance testing on 15+ dashboards to find issues with data load times and other database connection problems
- Collaborated with the IT intern team on a self guided project teaching company employees about Microsoft Power Automate
- Presented monthly to the IT leadership team about the internship progress

Student Help for Professor Yea-Seul Kim | 6/2022 - 6/2023

- Designed and conducted a user study focusing on low-vision populations
  - Examined a method from a previous paper for data visualization analysis in the context of alternative text
- Built HTML and Javascript prototypes with accessibility for blind populations in mind
  - Researched natural language processing methods to create a Python model and chatbot to assist blind populations with tabular data analysis

Projects

SemEval Task 8: Machine-Generated Text Detection | Fall 2023

- Explored the potential of simple traditional machine learning models (Random Forest, KNN, XGBoost, Logistic Regression) for binary classification of machine generated vs. human written test
- Used Python packages to extract lexical and syntactic features of the text to feed into models for training and evaluation
- Achieved test accuracy of over 0.80 on the training dataset, which was higher than the provided benchmark RoBERTa model
- Worked with LLMs like LLAMA and BERT to generate embeddings and other features to pass to the classification model

ECE561 Final Project: LLM-generated Synthetic Data Analysis | Fall 2024

- Interested in how well LLMs could generate realistic and diverse synthetic data in the medical field starting from an artificially unbalanced dataset
- Used both proprietary (ChatGPT 4o) and open-source (Llama 3.2, Gemma, Minstrel, Llama 3) LLMS and prompt engineering to generate synthetic data
    - Compared results of zero-shot and few-shot in-context learning
    - Analyzed datasets of many forms, including numeric and categorical tabular data and textual data
- Built baseline logistic regression classification models to evaluate the performance and quality of synthetic data

Involvement

Mentoring Chair for ACM-W (WACM) Chapter at UW-Madison | 9/2022 - 12/2024
Rewriting the Code (RTC) | 1/2023 - Present