Zirui Zhao

Shenzhen, China | ziruizhao0222@gmail.com | (+86) 15889599636

EDUCATION

- Bachelor of Engineering in Computer Science and Technology, Shenzhen University, China — Sep 2020 – Jul 2024 *(expected)*
- GPA: 4.08 / 4.5 (Top 4%) Average Score: 89.5/ 100
- Honors: "Liyuan Star" prize of Shenzhen University (the highest honor at the undergraduate level)
- Related Coursework: Data Structures, Probability Theory and Mathematical Statistics, Java Programming, Object-Oriented Programming, Algorithm Design and Analysis, Computer Systems, Database Systems, Big Data Processing and Analysis, Introduction to Machine Learning, Computer Network

SKILLS

- Software and Programming Languages: C++, Python, Java, MySQL, MATLAB
- Other Tools: Linux, Git, PyTorch, TensorFlow, Matplotlib, scikit-learn, Hugging Face

INTERNSHIP EXPERIENCE

Sangfor Technologies May 2023 - July 2023

Large Language Model Data Annotation Intern

- Applied security vulnerability scanning tools such as Fortify and AWVS to scan the source code to identify line numbers where error occurred and corresponding error types.
- Wrote automated Python scripts to split code segments into complete syntax blocks and fix the bugs.
- Deployed PyQt5 to develop a visual GUI platform for convenient data annotation, viewing and editing code.

PROJECT EXPERIENCE

Exploring Perceptual Experience of Complex Friction Patterns on Touch Surfaces Aug 2023 - Present

Professor Hasti Seifi, Arizona State University

- Utilized several NLP techniques, such as NLTK and the Spacy library in Python, to perform part-of-speech (POS) tagging on people's descriptions of different tactile signals.
- Implemented GPT-3.5 with prompt engineering and fine-tuning to extract 3 categories (emotional, sensational, and metaphorical) of keywords and phrases with an accuracy of 72%
- Leveraged static and dynamic word embeddings based on GloVe, ConceptNet and BERT to process vocabulary and employing hierarchical clustering to categorize the words into several groups.
- Evaluated the classification performance by Normalized Mutual Information (NMI) and calculated Pearson correlation to determine the relationship between signal features and words from each categories.

Open-Source Code Datasets Collection and Fine-Tuning Code Generation Models Jan 2022 - June 2022

Professor Qin Zhang(Shenzhen University) & Professor Meng Fang(University of Liverpool)

- Employed Python to integrate large code datasets such as Code Contest, CodeNet, etc., into a single large JSON file
- Created a web crawler using Python libraries (requests, requests_toolbelt, and html2text) to gather Leetcode coding questions categorized by different difficulty levels for future training purposes.
- Adopted ten-fold cross validation and fine-tune 7B LLaMA pre-trained language model to achieve code generation accuracy of 75% for given problems.

Emotion Recognition Inference System Based on Wearable Devices Sep 2021 - Dec 2021

Yongpan Zou, Shenzhen university

- Preprocessed the wearer's signals from the wearable devices using FFT, IFFT and a Butterworth band-pass filter.
- Compared LSTM, DFCNN, MobileNetV2, ResNet101, and DenseNet models and performed 5-fold cross-validation.
- Used DenseNet to categorize the wearer's emotions into 4 classes (happy, neutral, sadness, mixed), achieving precision and recall rates of 83.0% and 83.1%

Publication

- First Author, Wearable-Based Human Emotion Inference System, ICECI 2021: Edge Computing and IoT: Systems, Management and Security pp 128 – 135. Online ISBN：978-3-031-04231-7, Print ISBN：978-3-031-04230-0