

Joo Wan (Luke) Lim

10969 Wellworth Ave, Westwood Riviera, CA 90024
Tel: (424) 467-9693 Email: joowan980330@gmail.com

EDUCATION

UCLA Master of Engineering

Data Science

Cumulative GPA 3.68/4.0

Los Angeles, CA

Sept/2023-Present

Purdue University

Double Major in Data Science & Computer Science (Machine Intelligence Track)

Cumulative GPA 3.64/4.0

West Lafayette, IN

Sept/2016 – May/2023

WORK EXPERIENCE

Yonsei Softcomputing Lab

Summer Intern

Seoul, Korea

June/2023 – August/2023

- Contributed to research focused on enhancing Explainable AI (XAI) for anomaly detection in surveillance footage by integrating Visual Question Answering (VQA) methodologies. Involved generating insightful queries related to detected anomalies by an unsupervised neural network, aiming to provide clear and understandable explanations of anomalous scenes detected by the model
- Conducted comprehensive review of academic papers on Visual Question Answering (VQA) and Visual Anomaly Detection (VAD), synthesizing key findings and insights into detailed reports to inform and advance our research objectives
- Investigated essential features in surveillance footage and developed a question generation mechanism to formulate a series of questions using Python, adopting an approach analogous to the CLEVR dataset, to discern and elucidate the anomalous components within the scene

Guowoo International Co.

Summer Intern

Seoul, Korea

Summer 2022

- Developed a predictive model using Tensorflow and Keras that would output the probability of the buyer re-ordering the design of the clothes by analyzing how much each type of the design was ordered, the sizes, the color of the clothes, and the ingredients for the material that went into the production
- Examined financial data using Pandas and Numpy and informed the marketing team using data visualization tools such as Matplotlib and Seaborn to focus on specific designs of clothes and which buyers generating the most revenues
- Organized data about previous orders so that it is easily usable when training a machine learning model and created a format in Excel so that future data can be stored in the same form
- Built robotic process automation (RPA) for simple tasks for employees performing redundant tasks on Excel using Workbook and sent out advertisement email to potential corporate clients using smtplib, imap_tools, and account

ACADEMIC RESEARCH EXPERIENCE

Bag of Tricks to circumvent AI generated text detectors

Los Angeles, California

Project Leader

Winter 2024

- Executed a series of attacks on the AI vs HUMAN text dataset from Kaggle, utilizing techniques such as paraphrasing, random word removal/changing, and random typo insertion to challenge AI-generated text detectors
- Investigated the robustness of state-of-the-art AI-generated text detection models against various attack methodologies, quantitatively assessing their accuracy and reliability in distinguishing between human and AI-generated texts
- Identified vulnerabilities in existing detection methods through empirical analysis, providing insights to potential users and suggesting solutions to enhance the robustness and reliability of AI-generated text detection models

Bilingual SAS: Evaluating subset selection algorithms for contrastive self-supervised learning on text data

Los Angeles, California

Fall 2023

Project Leader

- Developed and implemented a bilingual subset selection algorithm for contrastive self-supervised learning, integrating both English and Spanish embeddings into the SAS framework to enhance text data representation
- Evaluated the performance of the original and bilingual SAS algorithms using AG News dataset, assessing downstream text classification accuracy with various augmentation techniques such as synonym replacement, random swapping, and GPT paraphrasing
- Conducted ablation studies to analyze the efficiency and usability of the SAS algorithm in the text domain, identifying key factors that influence the quality of learned embeddings and improving data efficiency by selecting optimal subset of unlabeled data

Word in Context NLP Project

West Lafayette, Indiana

Project Leader

Spring 2023

- Designed and implemented multiple PyTorch-based classifier, including deep averaging neural network, RNN, and LSTM, with various combinations of GloVe embeddings and neural architectures for identifying if the occurrences of a word in two contexts correspond to the same meaning or not
- Engineered new features by integrating part-of-speech information into the input layer of the neural network, resulting in 3% improvement in model performance
- Improved the performance of three models through meticulous hyperparameter tuning, modifying key parameters such as the learning rate, hidden size, number of layers, and non-linearity selection to achieve an accuracy rate of 62%

Indiana Family and Social Services Administration Capstone Project

West Lafayette, Indiana

Project Leader

Spring 2023

- Conducted a joint investigation with Indiana FSSA to perform an analysis of the relationship between urban/rural settings and availability of necessary healthcare and services
- Leveraged the scikit-learn library to develop SVM and decision tree models for predicting patient's urban/rural classification, achieving an accuracy rate of 79% and 69%, respectively
- Delivered a presentation on our findings to our peers and key stakeholders at FSSA, conveying the implications of our analysis in a clear and concise manner

Nuvve Data Mine Project

West Lafayette, Indiana

Undergraduate Data Science Researcher

Spring 2022

- Collaborated with Nuvve on implementing model that can forecast the occupancy rate of a charging station of electrical vehicles located across California and Europe to efficiently consume energy from the grid using V2G technology
- Pre-processed the data using Pandas, and Numpy so that we could train our model on the data
- Feature engineered new features based on weather conditions and day of the week to improve forecasting accuracy
- Built a forecasting model using PyTorch and Sklearn and improved the model accuracy by hyper-parameter optimization

My College Basketball Project

West Lafayette, Indiana

Project Leader

Fall 2021

- Designed relational schema for a webpage that displays College Basketball teams, players, and coaches
- Implemented the project using html, java, spring boot, and MySQL
- Utilized complex queries to retrieve the data that the user wanted to see from our database
- Employed different methods such as isolation level and indexing to fasten our queries so that the user does not need to wait for a long load time

Non-Authentic Review Analysis Project

West Lafayette, Indiana

Project Leader

Fall 2021

- Applied logistic regression model to classify whether a review of an online product was authentic
- Implemented the model using Scikit-learn library from Python
- Converted training data to matrix format using TF-IDF vectorizer and searched the parameters for the model
- Achieved accuracy score of 0.96774 out of 1.0 when evaluating the model against the testing data

EXTRA-CURRICULAR ACTIVITIES

Web Dev Club

West Lafayette, Indiana

Club Member

Nov/2021 - May/2023

- Acquired skills in web fundamentals such as HTML, CSS, Bootstrap, JavaScript, and NodeJS
- Built portions of personal website after every meeting
- Guided other non-computer science majors in learning basic coding skills required for web development

Competitive Programmers Union

West Lafayette, Indiana

Club Member

Nov/2021 - May/2023

- Developed skills needed to solve complex algorithms and data structures problems
- Stimulated growth in Purdue community's participation in ACM ICPC (ACM International Collegiate Programming Contest)
- Organized events for club members such as monthly codathon and movie nights

PKA (Purdue Korean Association)

West Lafayette, Indiana

Recruiting Manager

Fall 2019 - Fall 2020

- Established arrangements with large Korean Companies such as Samsung, LG, and SK to come visit Purdue University to provide company information to attract new employees
- Coordinated informational seminars for Korean Companies that were seeking to recruit students from Purdue University
- Organized different social networking events for the Korean community in Purdue such as Sports Day, and New Year's Day

COMMUNITY SERVICE

VocaliD

Online

Volunteer

Nov/2021 - May/2023

- Donated my voice to help develop personalized synthetic voices for people who are speech impaired for at least 4 hours per week
- Recruited new members to donate their voices to enable millions of people around the world who use the same computerized devices to communicate in their own individual unique voice
- Raised awareness of the lack of individuation of synthetic voices through social media

Smithsonian Digital Volunteers

Online

Volunteer

Nov/2021 - May/2023

- Transcribed more than 25 historical documents to typed documents that are searchable and machine-readable resources
- Transcribed over 100 rushed field notes, cursive handwriting, and faded text to be more readable to the future generations
- Reviewed and edited the work of other transcribers to ensure all transcriptions were accurate

Writing Wonders Organization

Online

Volunteer

Nov/2021 - May/2023

- Wrote digital letters for hospitalized children, elderly patients, and healthcare workers to provide words of encouragement to uplift their spirits for at least 4 hours per week
- Developed a web platform for writing supporting notes, letters, or cards to pediatric cancer patients at St. Jude Children's Research Hospital

HONORS AND AWARDS

- Dean's List *Spring 2023, Fall 2022*
- Dean's List & Semester Honors *Spring 2022*
- Dean's List & Semester Honors *Fall 2021, Spring 2021*
- Dean's List & Semester Honors *Fall 2021, Spring 2020*
- Dean's List *Fall 2020*
- Dean's List & Semester Honors *Fall 2019, Fall 2016*

SKILLS

- Programming in R and RStudio
- HTML, CSS, Javascript, Linux
- GGplot2, Data Frames, Matrices
- Python, NumPy, Pandas,
- Seaborn, Matplotlib, Plotly
- Scikit-Learn, Machine Learning, TensorFlow, PyTorch
- SAS, MySQL, MongoDB, Neo4j
- C, Java

LANGUAGES

- English (Fluent in reading, writing, speaking)
- Korean (Native in reading, writing, speaking)
- Mandarin (Fluent in reading, writing, speaking)

CERTIFICATES

- Udeyml Deep Learning: Advanced Computer Vision *July 2022*

- Udemy| PyTorch: Deep Learning and Artificial Intelligence *May 2022*
- Udemy| Python for Data Science and Machine Learning Bootcamp *May 2022*
- Udemy| R Programming A-Z: R For Data Science With Real Exercise *August 2021*
- Udemy| The Complete 2021 Web Development Boot Camp *July 2021*