NAIHAO XU

433 W Johnson Street, Apt 611, Madison, Wisconsin, 53703
+1 414 334 0163 ⋄ nxu43@wisc.edu

EDUCATION

University of Wisconsin-Madison WI Jan. 2024 - May 2025
Master of Computer Science
GPA: 4.0./4.0

University of Wisconsin-Madison WI Sep. 2019 - Dec. 2023
Bachelor of Science in Computer Science and Data Science
GPA: 3.661/4.0
Honor: Dean's List for five consecutive semesters
Research Interests: Large Language Model, Video Large Language Model, Natural Language Processing, Vision-Language Model, Human-robot Interaction, Human-computer Interaction,

RESEARCH EXPERIENCE

Face-Gemma: Emotion Analysis from Video Selfies University of Wisconsin-Madison
Supervisor: Prof. Vikas Singh — Research Student *May 2024 - Present*

- Developed Face-Gemma, a novel framework for analyzing emotions in video selfies by fusing facial landmark trajectories with language understanding, achieving 15% improvement over SOTA models.
- Created and curated a large-scale dataset of 70,000+ video selfies from 1500+ participants, capturing six ecological momentary assessment (EMA) dimensions for mental health monitoring.
- Implemented an innovative wavelet transform encoding for facial dynamics, processing 478 3D facial landmarks to capture subtle emotional expressions over time.
- Integrated MediaPipe Face Mesh for robust facial landmark tracking and designed a dual-stream architecture combining visual and verbal cues through prefix tuning with Gemma LLM.
- Demonstrated consistent model performance (MAE 0.79-0.82) across different demographic groups including age, gender, and race, ensuring fairness in emotional assessment.
- Conducted comprehensive ablation studies showing significant improvements through wavelet-based features (13.2% improvement) compared to traditional facial landmark approaches.

Bias on SOTA Multimodal Models and VLMs University of Wisconsin-Madison
Supervisor: Prof. Vikas Singh — Research Student *May 2024 - Present*

- Focus on identifying biases in state-of-the-art multimodal models and Vision Language Models (VLMs) across gender, age, and race during predictions.
- Evaluating models like Video LLaMA, Video LLaVA, ChatUnivi, and VideoGPT-plus on open-source datasets that contain gender, age, and race attributes, analyzing trends across these demographic factors.
- Fine-tuning Video LLaMA and LLaVA-NeXT using customized datasets to explore methods for reducing biases and improving fairness in predictions across diverse groups.
- Testing these models on a wide range of video, text, and audio inputs to assess how multimodal learning affects demographic-based biases, focusing on mitigating unfair predictions in various real-world applications.

Optimizing Data Proportions for Enhanced Model Performance University of Wisconsin-Madison
Supervisor: Prof. Frederic Sala — Research Student *Sep. 2024 - Present*

- Conducting research on training optimization to improve efficiency and effectiveness in selecting data proportions that optimize training and evaluation loss while preventing overfitting.
- Designing and running experiments on Spanish-to-English translation tasks using NeoGPT, focusing on reducing translation errors and enhancing model performance.
- Developing data-driven strategies for fine-tuning machine learning models, with an emphasis on balancing training efficiency and generalization capabilities.

Large Language Model Implementation on Home Robots University of Wisconsin-Madison
Supervisor: Prof. Bilge Mutlu — Research Student *July 2024 - Sep. 2024*

- Investigating to reduce the amount of hallucinations in most of the SOTA large language models while generating responses during human-robot interactions.
- Trying out different kinds of LLMs such as GPT-2 small, CLIP, T5 small, and BERT, and select the five models that only generate very few amounts of hallucinations by cross-comparing all usable models and the error thresholds we set.
- Propose a framework that utilizes multimodal GPT-4V to enhance embodied task planning through the combination of natural language instructions and robot visual perceptions.

Pest Identification Model Based on Multiscale CNN and ViT University of Wisconsin-Madison
Research Student *November 2023 - May 2024*

- Aim to develop a robust pest identification model using a combination of Multiscale Convolutional Neural Networks (CNN) and Vision Transformers (ViT) for enhanced accuracy and contextual understanding.
- Integrated cross-modal feature fusion by combining image data with textual descriptions to improve the model's capability to accurately identify pests in complex agricultural environments. Utilized Pyramid Squeezed Attention (PSA) mechanisms within the Contextual Transformer Network (CoTN) to efficiently manage multi-scale spatial information and focus on the most relevant features.
- Leveraged pre-trained models and transfer learning to accelerate the training process and enhance model scalability, making it adaptable to various types of pests and agricultural scenarios.
- Achieved a 15% increase in identification accuracy and a 20% improvement in reliability, demonstrating the model's effectiveness in real-world applications.

Research on Bias in Existing Facial Emotion Recognition Systems University of Wisconsin-Madison
Research Student *Jan. 2024 - Jun. 2024*

- Aimed to examine bias from current popular approaches and models such as VGG16, EfficientNet, and ImageNet, then calculate their accuracy, precision, Recall, and F1 respectively. Calculate the per-pixel squared error (summed across the three channels) using the mean pixel value of each image.
- Evaluate existing datasets including FER2013 and AffectNet used for training and testing purposes and in larger datasets with greater facial variation, fairness metrics generally remain constant, suggesting that racial balance by itself is insufficient to achieve parity in test performance across different racial groups.

Enhanced Traffic Object Detection Using Improved YOLOv8 University of Wisconsin-Madison
Research Student *January 2024 - July 2024*

- Helped develop YOLO-RC, an advanced traffic object detection model based on YOLOv8, optimized for accuracy in complex roadside environments.
- Integrated MBConv modules and the C3FB structure to enhance feature extraction and small object detection while reducing parameter count.
- Implemented a Bi-directional Feature Pyramid Network (BCFPN) for improved feature fusion and detection accuracy.
- Achieved a 4.5% improvement in mAP50 (91.1%) and a recall rate of 81.8% in real-world detection tasks.

Building Regression Methods for Global Temperature Prediction University of Wisconsin-Madison
Supervisor: Meenakshi Syamkumar — Research Leader *Sep. 2022 - Jun. 2023*

- Aimed to build multiple regression models including LN, SVM, and KNN, and utilize a unique dataset of existing climate model simulations to learn relationships between short-term and long-term temperature responses to different climate forcing scenarios.
- Evaluated and compared the effectiveness of models by calculating important numbers such as cross-validation. scores, mean squared errors, residuals of the predicted results, and $R2$ scores of the overall accuracy of the model.

INTERSHIP & TRAINEE

University of Wisconsin-Madison Madison, USA
Back-End Developer *Sep. 2021 - Dec. 2021*

- Built a music player together in a team as back-end a developer by using Java language to create testing schedules to optimize user interface and experience.
- Aimed to develop fast and efficient data acquisition from a database created by a data wrangler by utilizing the hash table and making it more efficient and effective for the front-end developer to use.
- Cooperated with the front-end developer to interact with users and provide services to them by building easy-to-use application interfaces with faster search.

University of Wisconsin-Madison Madison, USA
Data-Wrangler *Sep. 2021 - Dec. 2021*

- Built a CO2 emission level ranking project together in a team as a data wrangler.
- Aimed to mine raw CO2 emission level data from authoritative websites and read into csv/xlsx files.
- Further modified raw data by deleting irrelevant columns of data and specifying a special-character delimiter for the back-end developer to further store into data structures such as the hash table.

EXTRA-CIRRUCULAR

- International Teaching Assistant, Nanjing, China May 2019 - Aug. 2019

SKILLS AND INTERESTS

Interests	Tennis, Basketball, Large Language Model, Video Large Language Model, Vision-Language Model, Modeling and Simulation, Natural Language Processing Human-robot Interaction, Human-computer Interaction, Modeling Optimization
Computer	Java, Python (PyTorch, TensorFlow, NumPy, Scikit-learn, Matplotlib, SciPy, Pandas, StatsModels, Regressions, PyTorch, PyArrow), C, C++, R, Kafka, Docker, SQL, Spark, HDFS, HBase, Cassandra, BigQuery
Language	Mandarin (native), English (native)

PUBLICATIONS

Refereed Publications

- Yogesh Prabhu, Naihao Xu, Sotirios Panagiotis Chytas, Harshavardhan Adepu, Nicole Hendry, Matthew Kaharudin, Nathanael JK Vack, Ross Jacobucci, Simon B. Goldberg, Christine D. Wilson-Mendenhall, Richard Davidson, Vikas Singh. *Face-Gemma: Fusing Face Landmark tracks and Language from Video Selfies*. Submitted to the Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
- Naihao Xu, Hui Deng, Meijun Sun. *Algorithm for Detecting Traffic Objects in Complex Roadside Scenes Based on Improved YOLOv8*. IEEE Transactions on Circuits and Systems for Video Technology, 2024. Submitted; targeting SCI JCR Q1-ranked journal/conference. `https://github.com/lucasxu777/YOLOv8_traffic_object_detection`
- Naihao Xu, Hui Deng, Meijun Sun, and Zhiliang Qin. *Pest Identification Model Based on Multiscale CNN and ViT*. Computers and Electronics in Agriculture, 2024. Submitted; targeting SCI JCR Q1-ranked journal. `https://github.com/lucasxu777/Pest_Identification_CNN_ViT`