Yi-Shan (Annie) Wu

(659) 901-7448 | annie.wu.ysw@gmail.com | www.linkedin.com/in/anniewu-tw/ | www.anniewu-tw.com/

EDUCATION

SUNY-Binghamton University

PhD of Computer Science GPA: 4.00/4.00 *Dec 2025*

Poster*:* Detecting Overhead Obstacles for Robot Guide Dogs to Assist the Visually Impaired (Accepted NERC 2025)

- Built a ROS 2-based LiDAR perception pipeline on a Unitree quadruped, enabling real-time obstacle detection via point cloud decoding, plane removal (RANSAC), and DBSCAN clustering with spatial filtering; published 3D bounding boxes to RViz for visualization and analysis.
- Automated per-frame LiDAR logging to .pcd for post-experiment analysis and annotation.

Research*:* Facial Expression Recognition Using IMU Data (Expected Ubicomp)

- Designed and deployed a synchronized sensing platform on smart glasses using the ICM-20948 IMU and egocentric video for real-time expression monitoring.
- Built a cross-modal FER pipeline aligning temporal IMU signals (via AutonLab MOMENT pipeline) with video-based features extracted by ARBEx (IR50 + MobileFaceNet + ViT), using COMODO loss to project both modalities into a shared embedding space
- Enhanced model generalization through cross-validation on new user datasets.

EXPERIENCE

ectios

Data Entry Engineer Intern *Aug 2025 - Sep 2025*

- **Enabled solar adoption modeling across 10+ Spanish municipalities** by building unified, geocoded datasets (MITECO registry, INE census, DSO capacity), using Python ETL pipelines and AWS S3 data lake.
- **Improved data accessibility for incentive analysis by 40%** through **LLMaaS-based PDF parsing** and geocoding APIs, reducing manual extraction effort and ensuring schema consistency.
- **Accelerated cross-team integration** by documenting ingestion workflows and metadata standards, which streamlined collaboration between engineering and research groups across 3 time zones.

Gained international experience working in English with a cross-national team spanning Europe, Asia, and the Americas, ensuring smooth collaboration on data ingestion, processing, and modeling workflows.

PROJECTS

GAN-Based Data Augmentation & Model Fine-Tuning (https://shorturl.at/rBXrQ) *Jul 2024 - Dec 2024*

- Optimized adversarial data generation using GANs, cosine scheduler, and dynamic gradient adjustments to mitigate class imbalances.
- Increased model accuracy from 26% → 89% through fine-tuning and adaptive training strategies.
- Refined model robustness with hyperparameter tuning and Transformer-based fine-tuning for real-world deployment.

DrugPair2Vec – Biomedical Text Analysis (https://shorturl.at/7fZwv) *Jul 2023 - Dec 2023*

- Implemented PubMedBERT pretrained model on Colab GPU for biomedical text processing.
- Utilized DrugPair2Vec algorithm to analyze chemical interactions and drug relationships.
- Compared chemical similarities using Scikit-learn's Jaccard score and enriched findings with DrugBank Online data.

SKILLS

- **Programming Languages:** Python, C, Java, MySQL
- **AI/ML Libraries & Frameworks:** PyTorch, TensorFlow, LLM, Deep Learning
- **Algorithm & Optimization**: Data Structure, recommendation, parallel computing
- **Data Analytics & Visualization:** ETL pipeline, Power BI, Cloud Computing (AWS, GCP)
- **Certifications:** Project Management (PMP & CAPM)