Yinwei Zhang

☎ (+1) 520-486-7382 · ✉ zhangyinwei1994@gmail.com · ⌂ Tucson, Arizona (willing to relocate)
in https://www.linkedin.com/in/yinwei-zhang · ⚲ https://super-zyw.github.io/

PROFILE

Seeking MLE/DS/AS position. Highly motivated Ph.D. candidate with a strong passion for leveraging statistics, machine learning, and deep learning techniques to address real-world challenges. Skilled in Python and database programming languages. Hands-on industrial experience in computer vision, large language models, and MLOps.

EDUCATION

University of Arizona   *United States*
Ph.D. Candidate in Industrial Engineering   Aug, 2018 - Dec, 2023 (Expected)
Master of science in Statistics&Data Science   Aug, 2020 - Dec, 2022
Master of science in Engineering Management   Aug, 2016 - May, 2018

Jingchu University of Technology   *China*
Bachelor of science in Process Equipment and Control Engineering   Sep, 2012 - Jul, 2016

Related Coursework: Optimization, Probability&Statistics, Machine Learning, Database Design, Neural Networks

TECHNICAL PROFICIENCY

Programming skills: Python, R, Matlab, SQL, MongoDB
Frameworks: TensorFlow, PyTorch, scikit-learn, Numpy, Pandas, Langchain
Platforms: Git/Github, Docker, GCP, Jenkins, Kubeflow, MLflow, Airflow, FastAPI, Streamlit

EXPERIENCES

Machine Learning Engineer   May, 2023 - Aug, 2023
Intern, CVS Health   *Hartford, CT*
- Contributed to the development of the Python-based terminal user interface in **GCP Vertex AI**. This TUI streamlines the process of model building/deployment/management, boosting the efficiency of **MLOps**.
- Deployed **large language model** on the GCP endpoint using **custom containers** and **FastAPI**, allowing users to access ChatGPT-like services such as conversational chatbot and document Q&A. This reduces costs by **93%** while enabling **100%** compliance in data governance compared to calling external API.
- Evaluated LLM efficiency using various embedding methods, vector stores, and hardware. Demonstrated the instruction-finetuning of LLMs, which can improve performance in real-world business scenarios.

Research Scientist   Jan, 2020 - Aug, 2021
Intern, ABB Inc   *Raleigh, NC*
- Constructed supervised **deep learning pipelines** for computer vision applications in industrial settings using Python and its libraries, enabling accurate real-time predictions under complex scenarios.
- Enhanced pipeline efficiency through **fine-tuning models** with backbones and developed novel modules for data preprocessing and model evaluation. These led to a remarkable performance boost of **30%**.
- Elevated model performance by incorporating high-fidelity synthetic images during training, which addresses the overfitting that caused by limited real-world data and allows model evaluation in corner cases.
- Collaborated with team members to identify key evaluation metrics, write reports, and **file 3 patents**.

Research Assistant   Aug, 2018 - Present
University of Arizona   *Tucson, AZ*
- Developed a **Bayesian-based** matrix factorization method for object detection, delivering results comparable to fine-tuned deep learning methods in real-world benchmark datasets while reducing **99% memory usage**.
- Proposed a method to model the reliability of the **autonomous vehicle** system by designing a novel simulation pipeline. The paper is selected as the **best paper** finalist in ICQSR (the top conference in reliability).
- Proposed a **regularized regression** model for burst detection in a water distribution system. The basis functions are adopted for **dimension reduction** and an optimization algorithm is developed for model estimation.

PROJECTS

Burst Detection for Sensor System   *RNN, GCP, Streamlit, TensorFlow* (link)
Trained a Seq2Seq model on time-series signals in TensorFlow. Detected anomalies by analyzing the residuals between the predictions and observations. Served the model in GCP on demand by building a website via Streamlit.

Image-based Anomaly Segmentation   *Autoencoder, PyTorch* (link)
Trained an autoencoder using PyTorch. Identified abnormal features by the cosine similarity score and replaced them with top-k similar normal features. Segmented anomalies by model decomposition, achieving 0.88 in F1 score.

SELECTED PRESENTATIONS

- **Zhang, Y.** (2019). Tutorial: Applications of Spatial-Temporal Data Analytics in Industry. *Grand Lab Slam Workshop*, University of Arizona, Tucson
- **Zhang, Y.**, Liu, J., Lansey, K. (2019). Functional Data Analytics for Detecting Bursts in Water Distribution Systems. *INFORMS Annual Meeting*, Seattle
- **Zhang, Y.**, Liu, J., Son, Y. (2018). Effective and Efficient Moving Object Detection by a Moving Camera, *INFORMS Annual Meeting*, Phoenix

SELECTED PUBLICATIONS

- Xia, S., **Zhang, Y.**, Liu, J.. Context-Aware Bayesian Tensor Factorization. *Quality, Statistics, and Reliability Best Paper Competition*, INFORM Annual Meeting, 2023 (submitted).
- **Zhang, Y.**, Xia, S., Zhang, B., & Liu, J. (2023). Moving Object Detection from a UAV's Camera by Decomposing the Optical Flow. *The IEEE International Conference on Industrial Engineering and Engineering Management* (IEEM), Singapore, (submitted).
- Pan, F., **Zhang, Y.**, Liu, J., Head, L., Elli, M., & Alvarez, I. (2023). Reliability Modeling for Perception Systems in Autonomous Vehicles: a Recursive Event-Triggering Point Process. *INFORMS Conference on Quality, Statistics, and Reliability* (ICQSR), Raleigh, North Carolina (**best paper finalist**).
- Xia, S., **Zhang, Y.**, Zhang, B., Liu, J., (2023). Detecting Foreground in Videos via Posterior Regularized Robust Bayesian Tensor Factorization. *IEEE International Conference on Automation Science and Engineering* (CASE), Auckland.
- Pan, F., **Zhang, Y.**, Liu, J., Head, L., Elli, M., & Alvarez, I. (2022). Quantifying error propagation in multi-stage perception system of autonomous vehicles via physics-based Simulation. *Winter Simulation Conference*, Singapore.
- **Zhang, Y.**, Zhang, T., Liu, J., Kang, W., Liang, R., & Potter, B. (2022). Profile extraction for optical lens curing process with Image-based Regularized Tensor Decomposition. *Proceedings of the 2022 International Symposium on Flexible Automation* (ISFA), Japan.
- Nikravesh, Y., **Zhang, Y.**, Liu, J., & Frantziskonis, G.N (2022). A partition and microstructure-based method for large-scale topology optimization. *Mechanics of Materials*, Volume 166.
- Peterson, R. L., Shea, K. D., Liu, J., Luque, K., Powell, J., **Zhang, Y.**., Williams, D. K., Martin-Plank, L., Heasley, B. J., Phillips, L. R., & Crist, J. D. (2021). Family caregiving context: a pilot study. *The Arizona Nurse*, April.
- **Zhang, Y.**, Lansey, K., & Liu, J. (2020). Detecting bursts in water distribution system via penalized functional decomposition. *The IEEE International Conference on Industrial Engineering and Engineering Management* (IEEM) (**honorable mention award**).
- Lee, S., Jain, S., **Zhang, Y.**, Liu, J., & Son, Y. (2020). A multi-paradigm simulation for the implementation of digital twins in surveillance application. IISE Annual Conference.
- Lee, S., Jain, S., Yuan, Y., **Zhang, Y.**, Yang, H., Liu, J., & Son, Y. (2019). Design and development of a DDDAMS-based border surveillance system via UVs and hybrid simulations. *Expert Systems With Applications*, 109-123.

PATENTS

- **Zhang, Y.**, Zhang, Q., Zhang, B., Vidal, J., (2022). Robotic systems and methods used with installation of component parts, WO Patent WO2022265642A1
- **Zhang, Y.**, Zhang, Q., Zhang, B., Vidal, J., (2022). Robotic systems and methods used to update training of a neural network based upon neural networks outputs, WO Patent WO2022265643A1
- Zhang, Q., **Zhang, Y.**, Zhang, B., Vidal, J., (2022). System and method to generate augmented training data for neural network, WO Patent WO2022265644A1

LEADERSHIP/EXTRA-CURRICULAR ACTIVITIES

- **Organizer**, CVPR Data Challenge on Vision-based Industrial Inspection (link)   Jan, 2023 - Jun, 2023
- **President**, INFORMS Student Chapter at the University of Arizona   Aug, 2021 - May, 2023
- **Teaching assistant** for SIE 533, Fundamentals of Data Science for Engineers
- **Teaching assistant** for SIE 506, Quality Engineering