YICHEN ZHAO

https://shenshenzyc.github.io/yichen-zhao-portfolio/ | alexyczhao@gmail.com | linkedin.com/in/yichenzhao1

EDUCATION

University of Texas at Dallas	Ph.D. in Computer Science	*May 2026 (Exp.)*
University of San Francisco	M. S. in Data Science	*Jun 2023*
University of Pennsylvania	M. A. in Applied Mathematics and Computational Science	*May 2022*
Georgia Institute of Technology	B. S. in Mathematics, Applied Mathematics	*May 2020*

WORK EXPERIENCE

Data Science Intern, Dagshub Inc. – Developer Relations *Nov 2022 – Jun 2023*

Next Word Prediction

- Constructed a bidirectional LSTM network to predict subsequent words given text, harnessing MLflow's model tracking capabilities. Reduced cosine embedding loss by 57.60% compared to baseline model by choosing the top-performing experiment from logged experiments. (Dagshub repo)
- Coached Dagshub users on practicing MLOps in their ML projects by publishing a blog post on the MLOps-integrated workflow of deploying the next-word-prediction APIs in the Dagshub community. (Blog post)

Snowflake x Dagshub

- Forged integration between Dagshub tracking and a Snowflake database by leveraging Git to log SQL queries and capitalizing on DVC to log result tables. Tuned the SQL queries to slash MSE of the Lasso regression model that forecasts Lego set prices by 69.32%. Documented the ETL workflow in a blog post to enlighten Dagshub users on utilizing Dagshub tracking features when collaborating with cloud-hosted databases on Dagshub. (Dagshub repo)

Annotation Transfer for YOLO-formatted Annotations

- Automated the transition of YOLO-formatted annotations for rendering images with annotations in Dagshub repositories, slashing operation time by 80% versus open-source tools. (Colab notebook)

PROJECTS

Neural Style Transfer on Videos and Images (Github repo)

- Engineered a PyTorch-based program leveraging VGG19 model to execute neural style transfer, porting artistic style between images and videos by minimizing a loss function combining content and style loss. Optimized backpropagation process to reduce the stylized video generation time by 75%.

Hand Gesture to Sign Language Alphabet Translator (Github repo)

- Developed an interactive program using Python OpenCV library to capture live video and interpret hand gestures into sign language alphabet in real-time.
- Capitalized on Google's Mediapipe package to identify hand landmarks and fine-tune a random forest model to categorize their coordinates as sign language letters, achieving an accuracy of 92.52%.

ShenShenPL – A New Programming Language (Github repo)

- Designed ShenShenPL, an innovative programming language emphasizing readability and adaptability.
- Formulated intuitive syntax to enable programmers to author reusable and scalable code. Constructed an interactive shell to execute code, furnishing a user-friendly environment for prototyping and projects.

SKILLS

Python-related: Statsmodels, Pandas, Plotly, OpenCV, PyTorch, NLTK, SpaCy, Flask, Streamlit, BeautifulSoup, Airflow, PySpark, MLflow, Docker, Kubernetes, Metaflow, Evidently
Non-Python: MySQL, PostgreSQL, MongoDB, JavaScript, HTML/CSS, C/C++, Git, DVC
ML-related: Computer Vision, NLP, Deep Learning, Data Mining, MLOps, A/B Testing, Time Series Analysis