

Yi(Zoe) Zhu

Mobile: (530)2048986 | **Email:** yz656@duke.edu | **Website:** <https://zoe-zhu.netlify.app/>

LinkedIn: <https://www.linkedin.com/in/yi-zoe-zhu-625039164/> | **GitHub:** <https://github.com/zyzoe>

EDUCATION

Duke University, MS in Data Science (MIDS) (GPA: 3.9/4.0) **08/2020-05/2022**

Courses: Bayesian, Causal Inference, Deep/Machine Learning, NLP, Data Engineering, Algorithms, Cloud Computing

University of California, Davis, BS in Statistics & Psychology (GPA: 3.6/4.0) **09/2016-06/2020**

SKILLS

Certification: AWS Machine Learning Specialty, Neo4j Graph Data Science Certification Candidate

Programming: Python (scikit-learn, numpy, pandas, TensorFlow, Keras), R (Tidyverse, Shiny), SQL, Cypher

Tools: AWS, GCP (Google BigQuery), Azure, Tableau, Git, Flask, Neo4j, Google Analytics

EXPERIENCE

IEEE - Capstone Data Scientist **08/2021-Present**

- Leverage topic modeling and knowledge graph to enable efficient cross-disciplinary research and new concepts discovery for large-scale datasets (5.4M+ publications) hosted by IEEE *Xplore* digital library.
- Create automation to facilitate research collaboration across divisions and increase search efficiency.
- Design end-to-end ML pipelines to extract scientific concepts, validate through graph linkages; Tag paper with concepts based on relevancy, and establish the parent-child hierarchy between concepts.

Duke University - Graduate Teaching Assistant [\[site python\]](#) [\[site R\]](#) **05/2021-Present**

- Mentor graduate-level projects related to data wrangling, modeling, analysis, and application.
- Capture and solve coding problems in Python, R, Bash, Git for inter-disciplinary students.

MorphoSource Repository - Data Scientist Intern [\[demo\]](#) [\[poster\]](#) **05/2021-08/2021**

- Developed an interactive dashboard on self-defined metrics to visualize the impact of 3D scans of specimens. Derived insights to encourage user contribution and optimize data storage allocations.
- Remapped user information after database reconsolidation. Prototyped the data analysis pipeline with cleaning and processing.
- Worked and communicated across functions in a virtual environment setting to drive the product updates.

UC Davis FoxLab - Research Assistant **11/2018-06/2020**

- Implemented an advanced CNN architecture Deeplabcut for object detection with annotations to identify key body part coordinates for research subject video data.
- Performed clustering methods including PCA and tSNE on research subjects' behavior to classify their time-stamped emotional state. Contributed the analysis to the NIH grant proposal and received approval.

PROJECTS

A/B testing for Advertisement Tone Analysis **02/2021-04/2021**

- Designed and launched campaigns on LinkedIn for Duke University Data Science graduate program.
- Analyzed and evaluated the effect of tones on click through rate to improve the advertising strategies.

Flask Web App on AWS for Language Translation [\[github\]](#) **10/2020-11/2020**

- Devised a serverless data pipeline for language translation with continuous integration and deployment using AWS Services (CodeBuild, Translate, Elastic Beanstalk, etc.) and GitHub for version control.