

# Shawn H. Xu

[sx131@georgetown.edu](mailto:sx131@georgetown.edu) • 925-858-6938 • <https://www.linkedin.com/in/shawn-xu22>

San Ramon, CA, 94582

## Skills

**Programming Languages:** Python, JavaScript, SQL, R programming, MATLAB, HTML

**Tools:** Numpy, Pandas, Quarto, Shiny, Visual Studio, Jupyter Notebook, Microsoft Office Suite, Git/Github

**Languages:** English, Mandarin Chinese

## Work Experience

**LLMs Data Scientist – Data Annotation Tech**

*May 2024 – Present*

- Engineered code-related prompts to Large Language Models (LLMs).
- Ensured high standard of conversational AI performance by conducting detailed quality assurance on chatbot responses based on accuracy and reasoning.
- Techniques:** Python, Prompt engineering, Reinforcement Learning, LLMs, Error analysis

**Student Researcher/Intern – Dana Farber Cancer Inst./Harvard Medical School**

*May - July 2022*

- Studied structural functions of degraders through functional proteomics and genomics to profile synthetic degraders.
- Aided in the discovery of molecular glues that paved a new way in pharmacology to kill cancer cells.
- Co-authored a publication on Nature in relation to my work: <https://doi.org/10.1101/2023.02.14.528208>

## Projects

**Enhancing Skill Based Matchmaking Project - <https://shawnhxu.github.io/EnhancingSBMM/>**

- Programmed a classification model in predicting player ranks in the competitive video game *League of Legends* to determine the most impactful rank determining features in R and Python.
- Concluded an optimal direction for an improved Skill Based Matchmaking algorithm study.
- Techniques:** Naïve Bayesian Statistics, Exploratory Data Analysis, Data Cleaning, Feature Selection, Random Forest, Clustering, Scikit-Learn, Matplotlib

**Art Image Similarity Finder App - <https://github.com/5cminsublim/DSAN6600Proj>**

- Engineered a Streamlit App that takes in a user input image and outputs most similar art pieces based on image preprocessing steps.
- Techniques:** Python, Pytorch, Residual Networks (ResNets), Convolutional Neural Networks (CNNs), DeepLake, Streamlit, You Only Look Once (YOLO), Subject Segmentation, Image Embedding

**StarCraft2 Player Performance Analysis - <https://github.com/5cminsublim/DSAN5300Proj>**

- Conducted an in-depth analysis on in-game telemetry data and skill rating of StarCraft 2 players.
- Techniques:** Python, R Programming, ANOVA Testing, Pair-wise T-testing, Logistic Regression, Support Vector Machines (SVMs), Random Forest, XGBoost, SMOTE Balancing, Feature Clustering

**MNIST Classification Project – Supervised Machine Learning and Deep Learning**

- Developed Artificial Neural Network (ANN) model using the Modified National Institute of Standard and Technology (MNIST) dataset in Python.
- Evaluated a ~97% validation and test accuracy with final hyperparameter tuned ANN model.
- Techniques:** Data Preprocessing, L1/L2/Dropout regularizations, Grid Search, Early Stopping, Keras, TensorFlow, Scikit-Learn, Matplotlib, Multi-class Classification, Computer Vision

## Education

**Georgetown University, Washington, DC**

*August 2023 – Present*

- M.S. degree in Data Science and Analytics (DSAN). **GPA:** 3.95/4.0

**Boston University, Boston, MA**

*Sept 2019 - May 2023*

- B.S. degree in Biomedical Engineering. **GPA:** 3.5/4.0 – Dean's List of Academic Honor.