

Jianfei Xu

Data Analysis / Operations | Preferred city: Hefei | Email: andyxu1619@gmail.com

Applied Statistics undergraduate with experience in data workflows, AI annotation platforms, data cleaning, structured processing, Label Studio deployment, multimodal annotation, model output evaluation, and research documentation.

Education

Anhui University | B.S. in Applied Statistics | 2023 - 2027

- ? Relevant coursework: Time Series Analysis, Applied Stochastic Processes, Data Structures, Statistical Forecasting and Decision-Making, Multivariate Statistical Analysis, Java Programming.
- ? Participated in AI, data science, machine learning, and deep learning related training.
- ? Participated in a summer research program at the National University of Singapore.
- ? Published one ICCGV (EI) conference paper as a co-first author.

Experience

Research Assistant | Institute of Information Engineering, Chinese Academy of Sciences | 2025.10 - 2026.04

- ? Supported large-model-related data collection, cleaning, filtering, processing, and analysis.
- ? Participated in private deployment and Chinese localization of Label Studio for secure intranet research environments.
- ? Contributed to multimodal annotation, model output evaluation, deployment guides, and operation manuals.

Data Analysis Intern | iFlytek Co., Ltd. | 2024.07 - 2024.08

- ? Provided data analysis and quality assessment support for large-model-related tasks.
- ? Analyzed, filtered, cleaned, and organized model-generated data for training and evaluation workflows.
- ? Supported data quality checks and usability improvement with an accuracy level of approximately 80%.

Data Processing Intern | iFlytek Co., Ltd. | 2023.12 - 2024.02

- ? Handled data sorting, cleaning, classification, formatting, screening, deduplication, and structured processing.
- ? Assisted in checking model outputs, documenting data-related issues, and maintaining internal data records.

Projects

Million-Scale Medical Database End-to-End Construction | Data Engineer

- ? Worked on medical database construction for precision medicine research, including acquisition, cleaning, de-identification, structuring, version management, and quality control.
- ? Integrated electronic medical records, laboratory reports, and other sources; supported compliant processing and storage for over one million sensitive medical records.
- ? Used automated validation and missing-value handling strategies, achieving data completeness over 96% and annotation consistency above 94%.

Label Studio Private Deployment and Chinese Localization | Platform Lead

- ? Led Docker-based private deployment of Label Studio for an intranet research environment.
- ? Completed Chinese localization across interface text, prompts, error messages, and backend modules, with coverage above 98%.
- ? Prepared deployment and annotation documents; built a manual plus multi-model pre-annotation workflow related to one SCI paper under review.

Summer Research Program | Researcher, National University of Singapore

- ? Worked on local deployment of large models, visual question answering, and model output evaluation.
- ? Deployed Flamingo-3B, Flamingo-9B, Qwen-13B, and ChatGPT-MINI; wrote scripts for multi-dimensional scoring and analysis.

Skills

- ? Python, SQL, Java (basic), data cleaning, structured data processing, data quality validation, missing-value handling, field mapping.
- ? Million-scale data processing, database construction, data ingestion, version management, quality-control workflow design.
- ? Label Studio, Docker, model data processing, multimodal annotation, VQA, model output evaluation.
- ? Technical documentation, deployment manuals, annotation guides, research collaboration.