

# Haikuo Lu (Ezra)

Dual background in statistical modeling and system optimization, specializing in AI algorithm implementation and LLM engineering applications.

Experienced in nonlinear models, Agent frameworks, document parsing, and building high-performance systems with C++/Python for practical deployment.

## Education

### Boston University

**Master of Arts in Statistics** Boston, MA  
College of Math and Statistics GPA: 3.58/4.0  
2018.09 – 2020.01

- Core Courses: Statistical Learning, General Linear Regression, Bayesian Statistics, Computational Statistics, Stochastic Process

### Beijing Institute of Technology

**Bachelor of Science in Statistics** Beijing  
School of Mathematics and Statistics GPA 3.3/4.0  
2014.09 - 2018.06

- Core Courses: Mathematical Analysis, Advanced Algebra, Real Analysis, Probability Theory, Mathematical Statistics, C Programming

## Work Experience

### Bojia Pharmaceutical Technology Co., Ltd.

**Mid-level Algorithm Researcher** | Shanghai | 2023.08 – Present

#### Nonlinear Mixed Effects Model Algorithm Development:

- Reproduced cutting-edge paper algorithms (Paper → Python Demo → C++), achieving consistent results with industry gold-standard software
- Designed test cases, conducted data simulation and collection to validate algorithm accuracy
- Optimized computational performance by reducing redundant calculations, improving efficiency (e.g., 70%+ speedup by replacing Eigen with BLAS/LAPACK)

### Zhejiang Laiweilai Technology Co., Ltd.

**Data Scientist | Data Intelligence Department** | 2021.07 – 2023.08

- **Medical AI Project R&D:** Led AI algorithm design and implementation for medical research collaboration projects, covering data collection, model development, and algorithm deployment for hospital clinical research and management needs.
- **Data Collection and Cleaning Pipeline:** Built structured data pipelines for various data formats based on clinical department requirements, integrating patient data from different platforms.

- **Multimodal Data Processing:** Processed time series (e.g., heart rate waveforms) and medical images using ResNet and convolutional networks for feature extraction and fusion.

## Project Experience

### FOCE (First Order Condition Estimation) and SAEM Algorithm Development

- Led implementation from scratch of FOCEI and SAEM parameter estimation algorithms for Nonlinear Mixed Effects Models (NLME), including Hessian approximation, gradient computation, and optimizer integration.
- Optimized gradient solving with hybrid numerical and analytical methods based on cutting-edge literature, reducing iteration time by approximately 30%.
- Built reusable matrix and 3D array structures with memory pool management, reducing Eigen memory allocation bottlenecks and improving memory utilization in thousands of simulation tasks.
- Systematically compared outputs with commercial software (NONMEM, Monolix) across various model structures and real data simulations to verify accuracy and stability.
- Established complete modeling and evaluation workflow including model comparison, VPC prediction checks, and Bootstrap confidence interval estimation.

### LLM-based PDF Document Structured Information Extraction and Auto-labeling System

- **Background:** Large volumes of internal medical data stored as PDFs (clinical research protocols, statistical reports) requiring private LLM deployment due to data sensitivity. Statistics department frequently needed structured information extraction with strong automation demand.
- **System Design:** Built end-to-end pipeline from PDF to structured data using Dolphin and Deepseek R1, leveraging Dolphin for layout parsing and format standardization to output structured JSON.
- **LLM Labeling Interface:** Input structured JSON to privately deployed LLM, implementing entity recognition, field extraction, and label generation through prompt engineering and Agent architecture for automated data labeling.

- **Results:** System deployed in multiple clinical statistics projects, significantly reducing manual extraction workload and improving data extraction speed and accuracy.

### Personal Summary

---

- **Programming Languages:** Python, C++, R, Bash, SQL
- **Common Frameworks:** PyTorch, NumPy, Pan-

das, Polars, Scikit-learn, Eigen, SymPy, Open-BLAS/LAPACK

- **Modeling Methods:** FOCEI, SAEM, Bootstrap, Non-linear Mixed Effects Models, Bayesian Inference
- **Publication:** PAGE 31 (2023) Abstr 10676 [[www.page-meeting.org/?abstract=10676](http://www.page-meeting.org/?abstract=10676)]
- **Languages:** Chinese (Native), English (TOEFL 110)