

# Beichen Hu

28 E. Green Street., IL 61820 · (217) 954-3235 · [beichen7@illinois.edu](mailto:beichen7@illinois.edu) · [LinkedIn](#) · [GitHub](#)

## EDUCATION

---

### University of Illinois, Urbana-Champaign

Expected Graduation: May 2028

*Bachelor of Science in Statistics & Bachelor of Arts in Sociology*

Major GPA: 4.0/ 4.0

- Minor: Computer Science & Data Science
- Coursework: Machine Learning, R Programming, Python, Java, C++, Calculus, Discrete Structures, Statistics and Probability, Data Analysis, Data Visualization, Data Management, Social Research Methods

## PUBLICATIONS

---

Liu, H., Xu, Y., **Hu, B.**, & Dong, B. (2026). "The discerning principal: What prompting–uptake behaviors characterize high performers in human-AI collaboration?". *Computers & Education*. (Under Review)

**Hu, B.**, & Liu, H. (2025). "Intelligence Amplification: Cultivating Deep Thinking in the AI Era." *Shanghai Education: Global Education*, Issue 35, pp. 56-61.

## SKILL & CERTIFICATIONS

---

- **Functional:** Statistical Modeling, Data Cleaning, Data Visualization, Data Analytics, Machine Learning
- **Languages and tools:** Python, R, Java, C++, SQL, Stata, Microsoft Suite, GitHub

## PROFESSIONAL EXPERIENCE

---

### ATLAS – AI Exploration Team, University of Illinois Urbana-Champaign

Urbana, IL

*AI Research Intern*

Jan 2026 - Present

- Architecting a **Retrieval-Augmented Generation (RAG) prototype** to function as an academic advising assistant agent, designed to automate responses to logistical student inquiries using Python.
- Engineering a vector database pipeline to ingest and index unstructured administrative data (PDF handbooks, FAQs), enabling high-precision semantic search for complex degree requirements.
- Implementing robust prompt engineering and quantitative evaluation metrics to minimize hallucinations, benchmarking the model's accuracy and citation relevance against human advisor standards.

### Department of Statistics, University of Illinois Urbana-Champaign

Urbana, IL

*Course Assistant*

Aug 2025 - Present

- Supported instruction in an introductory statistics and data science course with 400+ students through labs, grading, and weekly office hours.
- Collaborated with instructors to enhance grading consistency and lab design.
- Guided students in **Python-based data analysis** (hypothesis testing, logistic regression, regularization, and machine learning).

### Professor Hui Liu's Research Team, Zhejiang University

Hangzhou, China/ Remote

*Research Assistant*

Mar 2024 – Present

- Analyzed unstructured human-AI interaction logs using Python and TF-IDF vectorization to transform sequential dialogue into high-dimensional feature sets.
- Identified distinct learner behavioral archetypes by implementing Hierarchical Clustering, uncovering latent patterns in how users transition between clarification and uptake.
- Engineered over 500 N-gram features and built a Random Forest classifier to predict performance, achieving statistical rigor through Leave-One-Out Cross-Validation (LOOCV) for small-sample optimization.
- Validated predictive behavioral sequences (e.g., *Clarify Standards -> Partial Uptake*) using Permutation Tests ( $p < 0.01$ ) to ensure the significance of feature importance.

### Glocal

*Community Leadership and Marketing Intern*

Remote

Aug 2024 - Aug 2025

- Supported the University Pilot Program across multiple institutions, including Priestley College (UK), by coordinating feedback and go-to-market (GTM) strategies
- Helped train an internal AI agent using pilot data to improve prediction accuracy and decision support.
- Promoted educational equity through cross-campus engagement and data-driven outreach.

## PROJECT

---

### **Sandia National Laboratories Data Challenge 2025 | *Python, Scikit-learn, Pandas* | Nov 2025**

- Developed a predictive machine learning pipeline to minimize scrap rates in Additive Manufacturing (AM) by analyzing high-dimensional build parameter data.
- Engineered features to quantify spatial risk, transforming complex coordinate data into Euclidean distance metrics, identifying positional variance as a key failure driver.
- Built an optimized Random Forest classifier using GridSearchCV to handle non-linear interactions between powder types and build layouts.
- Addressed class imbalance by implementing Probability Calibration (CalibratedClassifierCV), correcting model bias to produce reliable risk estimates.
- Conducted Precision-Recall threshold tuning, optimizing the decision boundary to 0.18 to achieve a 63% Recall, significantly reducing the risk of shipping non-conforming parts.
- Validated the technical feasibility of recycled powder usage via statistical testing (Kolmogorov-Smirnov Test) on particle size distributions.

## OTHER SKILLS & CERTIFICATIONS

---

- **Extracurricular Experience:** Champaign Photography Association (Spring 2025 - Present)
- **Awards & Certification:** Dean's list: Fall 2024, Spring 2025; Leadership Certificate (Aug 2025)
- **Language:** English (Fluent), Mandarin (Native)