

LUYANG SI

Champaign, IL | (+1) 447-802-1417 | ls36@illinois.edu | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

University of Illinois Urbana-Champaign, Champaign, IL 08/2022–08/2025
B.S. in Information Sciences + Data Science | GPA: 3.65/4.0
Relevant Courses: Data Management; Model & Learning in Data Science; Data Visualization; Statistical Analysis

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, Java, C++, Command Line
- **Data Analysis & Statistics:** pandas, scikit-learn, OLS regression, hypothesis testing, A/B testing, precision/recall, LDA, spaCy, BERT
- **BI & Visualization:** Power BI, Tableau, Matplotlib, KPI dashboard design, executive reporting
- **Data Engineering:** Azure SQL, medallion-layered pipelines (raw/cur/meta), watermarking, incremental ingestion, schema-drift detection

PROFESSIONAL EXPERIENCES

- Narciszmade**, Data Analyst/Information Consultant, Remote 11/2025–Present
- Design and maintain end-to-end sales analysis pipelines in Python and R; automated KPI reporting informs pricing and inventory strategy for executive review.
 - Architect and manage SQL databases ensuring data integrity, traceability, and cross-functional accessibility.
- Red then Black (RtB)**, Behavioral Research & Analytics Analyst, Remote 01/2026–Present
- Designed A/B-style calibration experiments and multi-round confidence-accuracy measurement sessions to surface signal reliability gaps; translated findings into structured product metrics for alpha-stage platform development.
 - Performed transcript-level semantic clustering and precision/recall analysis on decision sessions, converting behavioral patterns (confidence volatility and decision reversal) into executive-ready reports for founder-level strategy decisions.
- Information Quality Lab**, Research Assistant, Champaign, IL 08/2024–09/2025
- Built automated pipeline tracking indexing drift across 208 DOIs (retracted Apr 2023, unflagged Jul 2024); precision/recall analysis of retraction flag classifiers surfaced systemic data quality failures — co-authored manuscript under review + 2 published datasets.
 - Presented findings at 3 conferences (iSchool Research Showcase 2024; Undergraduate Research Symposium 2025; STEM Symposium 2025); supervised by Prof. Jodi Schneider.
- RCMoon Educational Organization**, Co-Founder & Vice President, Champaign, IL 03/2024–05/2025
- Co-founded student educational organization; collected and analyzed club social media database to surface growth insights and present data-driven recommendations to the board for content and outreach strategy.
- Business Intelligence Group**, Information Consultant, Champaign, IL 08/2024–12/2024
- Developed inventory demand forecasting model (Random Forest) on historical sales data + macroeconomic indicators (GDP, stock market, and import/export) to optimize stock management for a global food distributor.
 - Synthesized model outputs into data-driven operational recommendations; delivered executive briefings to client leadership driving procurement and replenishment strategy improvements.
- EasyTransfer**, Overseas Market Operator, Beijing, CN 05/2023–08/2023
- Analyzed user behavior data on financial products to design targeted advertising strategies; coordinated with student councils, banks, and financial agencies for overseas market outreach.

CERTIFICATION

- IBM Data Engineering Professional Certification
- Google Business Intelligence Professional Certification
- Google Project Management Professional Certification

PROJECTS

- Crossref Retraction Metadata Analysis** ([Link](#)) | Manuscript under review + 2 published datasets at UIUC Repository
- Automated anomaly detection pipeline tracking 208 DOIs with indexing drift; performed precision/recall analysis of retraction flag classifiers across journal titles — directly analogous to Trust & Safety signal reliability and data quality monitoring.
 - Built real-time pipeline collecting DOI metadata, extracting title keywords, and benchmarking against protocol flags; findings presented at three conferences (iSchool Showcase 2024; URS 2025; STEM Symposium 2025).
- Research Dataset Recommendation System** ([Link](#))
- NLP-powered tool matching research questions to 25+ data sources (CFPB, FRED, Census, WRDS), cutting discovery time from ~2 hrs to <5 min; includes end-to-end CFPB complaint analysis (OLS regression, t-tests) with reproducible code and publication-ready visualizations.
- Canvas Platform Data Ingestion** ([Link](#))
- Production-style medallion pipeline (raw/cur/meta) ingesting Canvas JSONL into Azure SQL with rerun-safe incremental ingestion via watermarking, schema-drift detection, and run-auditing metadata for data quality monitoring.
- Wikipedia Cross-Lingual Semantic Analysis** ([Link](#))
- Pipeline pairing EN ↔ ZH Wikipedia pages across sensitive identity categories (race, gender, nationality, religion); computed multilingual sentence-level semantic similarity using embeddings to surface low-similarity mismatch pairs for downstream content bias review.